P.1877/69

maszyny matematyczne

zastosowania

w gospodarce

technice

i nauce

3
1969

Zeszyt poświęcony problemom automatyzacji wyszukiwania informacji

SPIS TREŚCI

Str.

Dorota Prawdzic — „Struktura systemów informacyjnych i ich automatyzacja" 1

„Wprowadzenie do problemu wyszukiwania informacji" — Tłum. z ang. **J. G.** 3

Andrzej Makowski — „INBI — System przetwarzania na EMC informacji bibliograficznych" . 7

Tadeusz Markowski — „System IGA — Informacja Grupowana Automatycznie" 12

Władysław Klepacz, Tadeusz Markowski — „System KWOC — zastosowanie maszyny ZAM 41 do sporządzania indeksów bibliograficznych" 16

„Przegląd urządzeń do wyszukiwania i gromadzenia informacji" — Tłum. z ang. **J. G.** 19

СОДЕРЖАНИЕ

Д. Правдзиц — „Структура информационно-поисковых систем и их автоматизация"

— „Введение в проблему поиска информации"

А. Маковски — „INBI — система обработки библиографической информации на ЭВМ

Т. Марковски — „Система IGA — составление справочников на ЭВМ 1440

В. Клепач, Т. Марковски — „Система КВОК — применение ЭВМ ZAM-41 для составления библиографических указателей"

— „Обзор технических средств для поиска и хранения информации"

CONTENTS

D. Prawdzic — „The structure and automation of information retrieval systems"

Information retrieval: an introduction

A. Makowski — „INBI — system of reference information processing on ZAM 41 computer"

T. Markowski — „IGA system — IBM 1440 computer provided indexes

W. Klepacz, T. Markowski — „KWOC" — reference ZAM-41 computer produced indexes"

Survey of IS and R equipment

WYDAWNICTWA CZASOPISM TECHNICZNYCH NOT
Warszawa
Czackiego 3/5

KOLEGIUM REDAKCYJNE

Redaktor naczelny prof. dr Leon ŁUKASZEWICZ

Mgr inż. Jacek KARPIŃSKI, Władysław KLEPACZ, dr Antoni MAZURKIEWICZ, inż. Dorota PRAWDZIC (zast. redaktora naczelnego), mgr inż. Andrzej TARGOWSKI (zast. redaktora naczelnego)

Sekretarz Redakcji mgr Wanda KAČER **Redaktor techniczny** Alicja BIL

RADA PROGRAMOWA

Prof. dr inż. Jerzy Bromirski (przewodniczący), mgr inż. Jan Bursche, doc. Stefan Czarnecki, mgr Michał Doroszewicz, mgr Adam B. Empacher (sekretarz), mgr inż. Bolesław Gliksman, mgr inż. Józef Knysz, mgr inż. Ludwik Mebel, doc. dr Tadeusz Peche, inż. Zdzisław Puzdrakiewicz, doc. mgr inż. Józef Thierry (wiceprzewodniczący), dr Tadeusz Walczak, mgr Stefan Wojciechowski, dr inż. Henryk Woźniacki, mgr inż. Jan Z. Żydowo

Redakcja: Warszawa, ul. Emilii Plater 20 m. 15, tel. 21-13-91. Zastępca redaktora naczelnego tel. 28-37-29

Zakład Kolportażu WCT NOT, Warszawa, ul. Mazowiecka 12

Zakł. Graf. „Tamka". Z. 2. Zam. 44. Papier druk. powlekany V kl. 80 g. A-1. Obj. 3 ark. druk. Nakład 2250. P-98

Cena egzemplarza zł 8.— **Prenumerata roczna zł 96.00**

maszyny matematyczne

zastosowania w gospodarce, technice i nauce

Nr 3

MIESIĘCZNIK

1969

ROK V

Marzec

Organ Pełnomocnika Rządu do Spraw Elektronicznej Techniki Obliczeniowej i Naczelnej Organizacji Technicznej

POLITECHNIKA ŚLĄSKA BIBLIOTEKA GLIWICE

P.1877/69

Dorota PRAWDZIC
Instytut Maszyn Matematycznych
Warszawa

681.322.004.14:025.4:002.513.5

Struktura systemów informacyjnych i ich automatyzacja

Omówiono strukturę i problematykę systemów wyszukiwania informacji naukowo-technicznej i wskazano możliwości usprawnienia ich funkcji za pomocą automatyzacji.

Systemy informacji naukowo-technicznej obejmują następujące główne procesy:

1. Gromadzenie dokumentów i opracowywanie zbiorów
2. Przechowywanie dokumentów
3. Wyszukiwanie żądanych dokumentów i wiadomości
4. Udostępnianie informacji i dokumentów.

Systemy informacyjne, a głównie dobór dokumentów i organizacja zbiorów — powinny maksymalnie przewidywać potrzeby określonych grup użytkowników.

Z punktu widzenia udostępnienia informacji można wyróżnić następujące rodzaje systemów:
1. System przystosowany głównie do wydawania informacji **retrospektywnej** na poszczególne zapytania użytkowników

2. System udostępniający **bieżące** informacje wyselekcjonowane i adresowane do poszczególnych użytkowników

3. Systemy mieszane.

Najogólniej system wyszukiwania informacji — w tym również system zautomatyzowany — musi spełniać zadanie takiego opracowywania dokumentów, gromadzonych w zbiorze, aby w fazach następnych, tj. przy przetwarzaniu i udostępnianiu — można było bez czytania tych dokumentów wybrać ze zbioru najwłaściwsze informacje.

Przy dwustopniowej organizacji zbiorów, gdzie pierwszym stopniem są dokumenty pierwotne (książki, czasopisma, raporty itd.), a drugim — są opisy tych dokumentów (abstrakty, streszczenia, referaty, analizy dokumentacyjne, noty bibliograficzne lub tylko symbole klasyfikacyjne) — na wstępie, przy wprowadzaniu dokumentu do zbioru, przeprowadza się analizę każdego wprowadzanego dokumentu, z uwzględnieniem przydatności zawartych w nim informacji. Następnie sporządza się jego opis w wybranym języku informacyjnym czyli dokonuje się czynności **indeksowania.**

Opis ten powinien odwzorować podstawową treść dokumentu w sposób możliwie zwięzły, obiektywny i jednoznaczny.

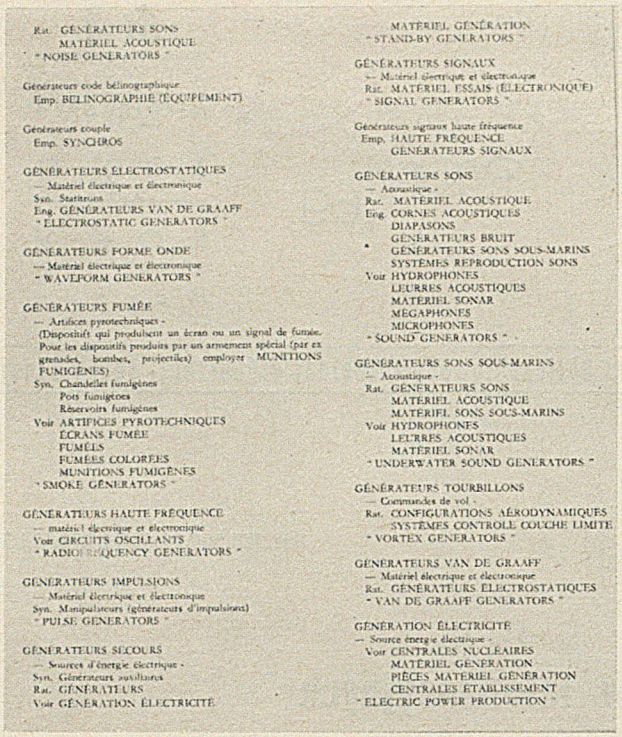
W ostatnich latach — w specjalistycznych systemach informacyjnych — rozwija się kierunek indeksowania dokumentów w deskryptorowych językach informacyjnych — za pomocą zestawu deskryptorów, tj. wyrazów najbardziej charakterystycznych dla treści tych dokumentów [1].

Jedną z głównych trudności w wyszukiwaniu informacji jest zapewnienie takiej komunikacji użytkownika z systemem informacyjnym, aby sformułowanie potrzeb informacyjnych użytkownika było dla systemu zrozumiałe i aby system wydawał użytkownikowi dokumenty ściśle odpowiadające jego zapytaniu. Powinna istnieć jak największa relewancja czyli — w naszym przypadku — zgodność znaczenia między zapytaniem użytkownika a odszukanym dokumentem z danego zbioru. Ponadto system powinien wydawać możliwie wyczerpujące informacje i eliminować informację zbędną.

Do tego celu służą **tezaurusy** czyli słowniki pojęć — podręczniki, którymi posługują się zarówno dokumentaliści, przy indeksowaniu nowych dokumentów, jak i użytkownicy, określając treść potrzebnych informacji w terminach deskryptorowego języka danego systemu.

Tezaurus (rys. 1) jest więc wspólnym podręcznikiem, pomagającym w porozumieniu się dokumentalisty i użytkownika z systemem informacyjnym. Powinien zawierać zestaw wszystkich potrzebnych deskryptorów łącznie z wyrazami bliskoznacznymi, które mogą występować w tekstach danej dziedziny i które można uważać za wyrazy kluczowe.

[1] Patrz: D. Prawdzic, A. Targowski — „Automatyzacja wyszukiwania informacji" — „Maszyny Matematyczne" nr 3/1967, s. 15—20.



Rat. GÉNÉRATEURS SONS
MATÉRIEL ACOUSTIQUE
" NOISE GENERATORS "

Générateurs code bélinographique
Emp. BELINOGRAPHIE (ÉQUIPEMENT)

Générateurs couple
Emp. SYNCHROS

GÉNÉRATEURS ÉLECTROSTATIQUES
— Matériel électrique et électronique
Syn. Statiteurs
Eng. GÉNÉRATEURS VAN DE GRAAFF
" ELECTROSTATIC GENERATORS "

GÉNÉRATEURS FORME ONDE
— Matériel électrique et électronique
" WAVEFORM GENERATORS "

GÉNÉRATEURS FUMÉE
— Artifices pyrotechniques -
(Dispositifs qui produisent un écran ou un signal de fumée. Pour les dispositifs produits par un armement spécial (par ex grenades, bombes, projectiles) employer MUNITIONS FUMIGÈNES)
Syn. Chandelles fumigènes
Pots fumigènes
Réservoirs fumigènes
Voir ARTIFICES PYROTECHNIQUES
ÉCRANS FUMÉE
FUMÉES
FUMÉES COLORÉES
MUNITIONS FUMIGÈNES
" SMOKE GENERATORS "

GÉNÉRATEURS HAUTE FRÉQUENCE
— matériel électrique et électronique
Voir CIRCUITS OSCILLANTS
" RADIOFREQUENCY GENERATORS "

GÉNÉRATEURS IMPULSIONS
— Matériel électrique et électronique
Syn. Manipulateurs (générateurs d'impulsions)
" PULSE GENERATORS "

GÉNÉRATEURS SECOURS
— Sources d'énergie électrique -
Syn. Générateurs auxiliaires
Rat. GÉNÉRATEURS
Voir GÉNÉRATION ÉLECTRICITÉ

MATÉRIEL GÉNÉRATION
" STAND-BY GENERATORS "

GÉNÉRATEURS SIGNAUX
— Matériel électrique et électronique
Rat. MATÉRIEL ESSAIS (ÉLECTRONIQUE)
" SIGNAL GENERATORS "

Générateurs signaux haute fréquence
Emp. HAUTE FRÉQUENCE
GÉNÉRATEURS SIGNAUX

GÉNÉRATEURS SONS
— Acoustique -
Rat. MATÉRIEL ACOUSTIQUE
Eng. CORNES ACOUSTIQUES
DIAPASONS
GÉNÉRATEURS BRUIT
GÉNÉRATEURS SONS SOUS-MARINS
SYSTÈMES REPRODUCTION SONS
Voir HYDROPHONES
LEURRES ACOUSTIQUES
MATÉRIEL SONAR
MÉGAPHONES
MICROPHONES
" SOUND GENERATORS "

GÉNÉRATEURS SONS SOUS-MARINS
— Acoustique -
Rat. GÉNÉRATEURS SONS
MATÉRIEL ACOUSTIQUE
MATÉRIEL SONS SOUS-MARINS
Voir HYDROPHONES
LEURRES ACOUSTIQUES
MATÉRIEL SONAR
" UNDERWATER SOUND GENERATORS "

GÉNÉRATEURS TOURBILLONS
— Commandes de vol -
Rat. CONFIGURATIONS AÉRODYNAMIQUES
SYSTÈMES CONTRÔLE COUCHE LIMITE
" VORTEX GENERATORS "

GÉNÉRATEURS VAN DE GRAAFF
— Matériel électrique et électronique
Rat. GÉNÉRATEURS ÉLECTROSTATIQUES
" VAN DE GRAAFF GENERATORS "

GÉNÉRATION ÉLECTRICITÉ
— Source énergie électrique -
Voir CENTRALES NUCLÉAIRES
MATÉRIEL GÉNÉRATION
PIÈCES MATÉRIEL GÉNÉRATION
CENTRALES ÉTABLISSEMENT
" ELECTRIC POWER PRODUCTION "

Rys. 1. Przykład strony tezaurusa (według „GESTION", nr 11/68)

W tezaurusie powinny być odzwierciedlone tzw. stosunki paradygmatyczne, m.in.: synonimii, hierarchii, współpodrzędności, funkcjonalnego podobieństwa, przyczyny—skutku, części—całości [2].

W tezaurusie powinny być również podane reguły tłumaczenia wyrazów kluczowych języka naturalnego na deskryptorowy język informacyjny.

Sporządzenie tezaurusa informacyjnego dla jakiejś dziedziny wymaga zwykle kilku lat pracy. Szczególnie jest to trudne w przypadku nowej dziedziny nauki lub techniki, w której terminologia nie jest ustalona, mało jest literatury dziedzinowej i brak słowników.

Wprowadzenie języków deskryptorowych jako języków systemów informacyjnych jest obecnie uważane za punkt wyjścia do automatyzacji podstawowych procesów informacyjnych. Języki deskryptorowe umożliwiają głębszą analizę dokumentów, bardziej szczegółowy ich zapis, lepsze kojarzenie. Są układami otwartymi, pozwalającymi na stosunkowo łatwe wprowadzanie nowych pojęć. Stwarzają pomost pomiędzy rozwojem różnych dziedzin nauki a rozwojem teorii i praktyki informacji naukowo-technicznej.

Rozwój języków deskryptorowych jest ściśle związany z ustawicznym wzbogacaniem terminologii specjalistycznych; musi uwzględniać zarówno normalizację terminologiczną, jak i tzw. dialekty dziedzinowe, a nawet środowiskowe. Stąd widać, jak ważne jest dostatecznie wczesne ustalanie terminologii.

Czynności indeksowania, tj. sporządzania w języku informacyjnym — na podstawie przeprowadzonej analizy — krótkiego opisu zawartości treściowej dokumentu wprowadzanego do zbioru nie udaje się jeszcze w pełni zautomatyzować w sposób zadowalający. Dotychczas tylko człowiek — i to o dużych kwalifikacjach — może gruntownie przeanalizować dokument, ocenić jego wartość informacyjną oraz ewentualne dziedziny zastosowania zawartej w nim informacji. Tylko człowiek może określić punkt wyjściowy do ewentualnego późniejszego odszukania tego dokumentu.

Wykwalifikowany dokumentalista — specjalista danej dziedziny — może prawidłowo i wystarczająco opisać zawartość dokumentu, nadając mu symbole klasyfikacyjne lub dobierając właściwe deskryptory.

Deskryptory wybiera się bezpośrednio z tekstu dokumentu lub przez skojarzenie wyrazów kluczowych występujących w tekście dokumentu z odpowiednimi wyrazami zawartymi w tezaurusie.

Czynność porównywania wybranych wyrazów kluczowych z deskryptorami można już wykonywać automatycznie.

W praktyce dotychczasowej powszechnie stosuje się **najprostszą metodę automatycznego indeksowania,** mianowicie metodę indeksów permutacyjnych (rys. 2), np. KWIC lub KWOC [3]. Polega ona w zasadzie na maszynowym wybieraniu deskryptorów **z tytułów** dokumentów. W przypadku gdy tytuły są mało znaczące, do powyższych czynności automatycznych muszą dojść pewne czynności człowieka, mianowicie analiza samego dokumentu i uzupełnienie tytułu dodatkowymi deskryptorami.

Bardziej rozwiniętą metodą automatycznego indeksowania jest maszynowe wybieranie deskryptorów **z analiz dokumentacyjnych** (streszczeń, abstraktów, referatów) wykonywanych uprzednio przez człowieka. Problem automatycznego sporządzania abstraktów znajduje się dopiero w stadium prób i jest chyba jeszcze dość daleki od rozwiązania. Niektóre proponowane metody opierają się na zasadach statystycznych, np. na ocenie częstotliwości występowania wyrazów w tekście.

Do pełnego zautomatyzowania obecnie nadają się więc tylko niektóre czynności występujące w procesach informacyjnych jak np.: **porządkowanie** (sortowanie), **wybieranie, porównywanie, drukowanie.** Przy uproszczonym indeksowaniu automatycznym (metoda indeksów permutacyjnych) otrzymujemy wydawnictwo zbioru uporządkowanego, w szczególności według kryterium alfabetycznego. Taki system zautomatyzowany może wydawać też indeksy autorskie i inne.

Zbiór uporządkowany według określonych kryteriów ułatwia wybieranie informacji z systemu każdorazowo na żądanie, zgodnie z zapytaniem wyrażonym w języku informacyjnym. Pytanie to często trzeba przetłumaczyć z języka naturalnego na język systemu, a czynności translacji takich zapytań mogą być zautomatyzowane. Wkraczamy tutaj w dziedzinę **przekładu automatycznego,** związanego z problemami semantyki i gramatyki. W celu rozwiązania tych problemów należy opracować algorytmy analizy i syntezy semantycznej, składniowej i morfologicznej. Prace te są dość szeroko na świecie prowadzone, ale pełnego rozwiązania jeszcze nie znalazły.

Przy indywidualnym udzielaniu informacji retrospektywnej na zapytania można wprowadzić formę dialogu między użytkownikiem i systemem, polegającym na iteracyjnym zbliżaniu się do coraz bardziej dokładnego merytorycznego doboru informacji. Przy zastosowaniu transmisji danych dialog ten może być również prowadzony na odległość.

W systemie informacji adresowanej tego rodzaju dialog jest utrudniony. Powstaje zatem problem takiego sprecyzowania zakresu zainteresowań użytkownika, które by dla systemu były możliwie jednoznaczne. Powiększanie liczby deskryptorów może nie dawać jeszcze odpowiedniego rozwiązania. Niektóre rozwiązania idą w kierunku nadawania deskryptorom określonych współczynników wagowych. Wprowadza się również metody etapowego określania relewancji. Dalsze rozwiązania prowadzą do stosowania operacji logicznych i modelowania.

---

2) Patrz: A. I. Czernyj — „Obszczaja mietodika postrojenja tiezaurusow", Nauczno-Tiechniczeskaja Informacja, Seria 2, nr 5/1968, s. 9—32.

3) Patrz artykuł: W. Klepacz, T. Markowski — „System KWOC" nr bieżący, s. 16.

Rys. 2. Przykład strony indeksu KWIC (według „American Documentation" nr 1/68)

SHAW SEISMOGRAPH	LOW MAGNIFICATION ATTACHMENT FOR A MILNE-	PETEJH-39-0341
SOURCE IN A CYLINDRICAL	LOW VELOCITY MEDIUM / AZIMUTHAL ASYMMETRY	MEECWC-62-0139
THE LAYER OF RELATIVELY	LOW WAVE VELOCITY AT A DEPTH OF ABOUT 80	GUTEB -48-0121
ITH VISCOUS COUPLING/ A	LOW-COST VERTICAL-COMPONENT SEISMOGRAPH W	MCFAJ -41-0139
H-WAVE EVIDENCE FOR THE	LOW-VELOCITY ZONE IN THE MANTLE / RAYLEIG	TAKEH -59-0355
S)	LUMINOUS PHENOMENA OF EARTHQUAKES (LIGHT	BALLCM-13-0187
SOUTHERN CALIFORN/ THE	LUNAR TRIGGERING EFFECT ON EARTHQUAKES IN	ALLEMW-36-0147
PERSONAL RECORD OF ADA	M. TROTTER OF CERTAIN AFTERSHOCKS OF THE	LOUDGD-44-0199
ROGRAMS BY MEANS OF THE	M.I.T. DIFFERENTIAL ANALYZER / DETERMINAT	RUGEAC-43-0061
RTH CALCULATED FROM THE	MACELWANE P CURVE, 1933 / VELOCITY OF P W	DAHMCG-36-0001
ERIMENTS WITH A SHAKING	MACHINE / EXPERIMENTS WITH A SHAKING MACH	ROGEFJ-50-0147
S DIRECT FROM A SHAD/ A	MACHINE FOR REPRODUCING EARTHQUAKE MOTION	RUGEAC-36-0201
INTO FORCED VIBRATI/ A	MACHINE FOR SETTING STRUCTURES AND GROUND	BLUMJA-35-0361
DATA (COMPUTER)	MACHINE PROCESSING OF SEISMIC TRAVEL-TIME	BOLTBA-61-0259
ION OF FOCAL DEPTH FROM	MACROSEISMIC DATA / ON THE ESTIMATION OF	BLAKA -41-0225
ION OF FOCAL DEPTH FROM	MACROSEISMIC DATA / ON THE ESTIMATION OF	NEUBH -38-0259

Postać wydawnicza wyników przetwarzania w zautomatyzowanych systemach wyszukiwania informacji może być dostosowana do bezpośredniego celu. Wyniki drukowane mogą służyć jako matryce do powielania, przy tym jest możliwe sterowanie programowe niektórymi procesami wchodzącymi w zakres redakcji technicznej. Według zapisu wyników można reprodukować karty dokumentacyjne w celu rozsyłania do abonentów. Można przesyłać zapis wynikowy na różnych nośnikach, m.in. na taśmach magnetycznych. Najnowocześniejsze rozwiązania przewidują zdalne przesyłanie informacji do odbioru na ekranach u użytkownika. System elektronicznego przetwarzania może być sprzężony z wyborem mikrofilmów i bezpośrednim odczytem pełnej treści żądanego dokumentu.

Największą jednak obecnie trudność dla automatyzacji systemów wyszukiwania informacji naukowo-technicznej stanowi rozwiązanie **problemów podstawowych** związanych z teorią informacji naukowej i zazębiających się z wieloma kierunkami badań w takich dziedzinach, jak semiotyka (ogólna teoria systemów znakowych), rozpoznawanie obrazów, strategia wyszukiwania, teoria wielkich systemów, lingwistyka matematyczna, języki sformalizowane, modelowanie matematyczne i inne.

Należy stwierdzić, że obok problemów komunikacji „człowiek—człowiek" „człowiek—maszyna", powstaje nowy problem: komunikacja **„człowiek nr 1—maszyna—człowiek nr 2"**. Porozumienie się człowieka z człowiekiem stwarza wiadome trudności. Włączenie do tej pętli maszyny — znacznie te trudności potęguje.

681.322.004.14:002.513.5

Wprowadzenie do problemu wyszukiwania informacji[1)]

Skrócone tłumaczenie artykułu R. M. Hayes, „Datamation", marzec 1968, nr 3. W artykule zastosowano podział systemów wyszukiwania informacji na trzy rodzaje: systemy oparte na zbiorach danych podstawowych, systemy referencyjne (bibliograficzne) i systemy przetwarzania tekstów. Omówiono problemy: użyteczności systemów, komunikacji, określania charakterystyk wyszukiwawczych oraz organizacji zbiorów.

Wyszukiwanie informacji nie jest — oczywiście — pojęciem nowym; stanowi ono integralną część procesu komunikacji. Informacje naukowe zapisywano od wieków i od dawna też dostępne były różnego rodzaju metody ich gromadzenia i wyszukiwania. W ostatnich latach niektóre problemy z zakresu wyszukiwania informacji przerosły możliwości dawnych środków technicznych, lecz pojawiły się obecnie nowe urządzenia do przetwarzania danych. Zadaniem nowoczesnego systemu wyszukiwania informacji jest wykorzystanie do rozwiązywania tych problemów — zdobyczy ostatnich 20 lat w dziedzinie rozwoju maszyn matematycznych, procesów fotograficznych i techniki zapisu magnetycznego.

Problemy wyszukiwania informacji wymagające — jak się wydaje — stosowania tego rodzaju środków technicznych wzrastają w nader szybkim tempie. Wskazywano już niejednokrotnie, że wobec ogromnej ilości powstających i rozpowszechnianych dziś informacji wysiłek ludzki może okazać się w wielu przypadkach daremny tylko dlatego, że nie ma możliwości ustalenia, czego już w danej dziedzinie dokonano. Sam ogrom informacji byłby może dostatecznym powodem uzasadniającym konieczność nowego podejścia do tej sprawy. Jeszcze ważniejszy jest chyba fakt, że poszczególne „cząstki" informacji pozostają zamknięte w ramach coraz bardziej ograniczonych obszarów specjalizacji, przy jednoczesnym znacznym prawdopodobieństwie, że te same informacje mogą być przydatne również dla innych obszarów specjalizacji. Rozwój specjalizacji jest powszechnie znanym zjawiskiem, rzutującym w sposób istotny na problemy kompleksowego podejmowania decyzji. Specjalizacja prowadzi jednak do powstawania specyficznej terminologii i żargonu, do ograniczenia obszarów bezpośredniej komunikacji oraz do utrudnienia dostępu do różnych wąskich dziedzin. Chociaż informacje z jednej dziedziny mogą być potrzebne do podjęcia decyzji w jakiejś innej dziedzinie, to niemal nieprzekraczalną barierę, uniemożliwiającą rozpoznanie tego związku, stanowi odrębność tematyki i języka.

Zagadnienie to można zilustrować szeregiem przykładów z dziedziny nauki i techniki. Powszechnie znany jest proces specjalizacji dyscyplin naukowych. Wielorakie zastosowanie informacji jest szczególnie widoczne w rozwoju nauk międzydziedzinowych.

Zagadnienia specjalizacji i różnorodnego wykorzystywania tych samych informacji występują również w dziedzinie gospodarowania i rządzenia. Tak więc — mimo że prace nad wyszukiwaniem informacji uzasadnione były głównie potrzebami nauki i techniki, to także i inne dziedziny w znacznej mierze będą z nich w podobny sposób korzystać.

---

1) Skrót artykułu zamieszczonego w amerykańskim czasopiśmie „Datamation", marzec 1968 r., nr 3, s. 22—26. Autor — dr nauk matematycznych Robert M. Hayes — jest dyrektorem instytutu badań w dziedzinie bibliotekarstwa Uniwersytetu California oraz profesorem szkoły bibliotekarskiej UCLA.

Termin „wyszukiwanie informacji" nabrał tak szerokiego znaczenia, że wydaje się celowe dokonanie podziału różnych rodzajów systemów na trzy grupy: systemy oparte na zbiorach danych podstawowych, systemy referencyjne (bibliograficzne) i systemy przetwarzania tekstów.

Systemy zbiorów danych podstawowych. Najbardziej rozpowszechnione wśród systemów wyszukiwania informacji są systemy operowania kartotekami danych podstawowych, w których zapis jest podzielony na określone rodzaje danych. Schemat takiego zapisu przedstawiony jest w tablicy I.

Zbiór tych zapisów może być uporządkowany według jakiegoś rodzaju danych (np. chronologicznie według dat urodzenia) i nazywa się kartoteką.

Tablica I
Schemat zapisu w kartotece danych podstawowych

rodzaj danych	rodzaj 1	rodzaj 2	..... rodzaj n
zapis	dane 1	dane 2	..... dane n
Przykład zapisu danych personalnych	Imię i nazwisko	Nr ubezpieczeń społecznych	Data urodzenia
	J. Smith	123-45-6789	12.03.26

W ostatnich latach pojawił się szereg „uogólnionych programów operowania kartotekami". Opisy niektórych takich programów ukazały się w „Datamation", numer ze stycznia 1968 r.

Programy te umożliwiają uzupełnianie kartotek dodatkowymi zapisami oraz zmiany w treści poszczególnych istniejących zapisów. Pod tym względem są one podobne do typowych systemów przetwarzania danych. Mają one jednak większe znaczenie dla celów sekwencyjnego wyszukiwania informacji, a jednocześnie umożliwiają przeszukiwanie kartoteki celem uzyskiwania odpowiedzi na zapytania wyrażane w postaci kombinacji Boole'owskich używających stwierdzeń logicznych *i, lub* oraz *nie*. Takim zapytaniem może być np.: „Znaleźć wszystkie zapisy zawierające informacje „A" z rodzaju danych 1 i „B" z rodzaju 2 lub „C" z rodzaju 1 i „D" z rodzaju danych 3.

Przykład systemu opartego na zbiorze danych podstawowych można znaleźć w przetwarzaniu danych gospodarczych, dostarczanych w dużej liczbie przez codzienną działalność urzędniczą. Systemy informacji do celów zarządzania stosuje się do identyfikowania i gromadzenia tych danych oraz do ich wyszukiwania w celu dalszego wykorzystywania, zwłaszcza do podejmowania decyzji. Inny przykład stanowią wielkie banki danych, które mają za zadanie zbieranie danych powstających w ramach bieżących operacji agend rządowych, ewidencjonowanie tych danych, wyszukiwanie oraz analizowanie.

Inne przykłady — to szpitale i ośrodki zdrowia (m. in. Szpital Ogólny w Massachussets oraz Zarząd d.s. Weteranów), które zakładają zbiorcze systemy zapisów medycznych. Dane pochodzą z działalności administracyjnej: rejestry przyjęć, wyniki różnych badań, zalecanie lekarstw itp., kontrolowanie *„on-line"* stanu pacjentów oraz z bezpośrednich zapisów dokonywanych przez lekarzy i pielęgniarki. Wykorzystuje się je do celów diagnostyki, opieki nad ciężko chorymi pacjentami, prac badawczych oraz administracji szpitalnej. Najbardziej chyba rozwiniętymi systemami są systemy dowodzenia i kontroli wykorzystywane do celów operacji wojskowych. Na przykład system SAGE prowadzi nasłuch i rejestrację sygnałów radarowych i telemetrycznych, porównuje je z uprzednio uzyskanymi danymi, gromadzi i przechowuje rezultaty, wyszukuje i udostępnia je w razie potrzeby analistom oraz przekazuje decyzje punktom dowodzenia w zakresie obrony kontynentalnej.

Tablica II
Systemy bibliograficzne

Instytucja	Zastosowanie systemu bibliograficznego
American Bibliographical Center Amer. Ośrodek Bibliograficzny	do sporządzania indeksów abstraktów z dziedziny historii
American Petroleum Institute — Division of Refining Amerykański Instytut Naftowy Oddział Rafinerii	dla potrzeb centralnej służby sporządzania abstraktów i indeksów
American Society for Metals Amerykańskie Towarzystwo d.s. Metali	dla serwisu dokumentacyjnego
Applied Mechanics Review Przegląd Mechaniki Stosowanej	do sporządzania indeksów wyrazowych i autorskich
Atomic Energy Commission Komisja Energii Atomowej	do sporządzania abstraktów z dziedziny atomistyki
Bio-Science Information Service Serwis informacyjny nauk biologicznych	do sporządzania abstraktów z dziedziny biologii
R. R. Bowker Company	do sporządzania publikacji tygodniowej, BPR
Chemical Abstracts Service Agencja Abstraktów Chemicznych	do wydawnictw: Chemical Titles, Chemical Biological Activities, Registry System
Clearinghouse for Federal Scientific and Technical Information Federalny Ośrodek Informacji Naukowo-Technicznej	dla dostarczenia informacji naukowo-technicznej w skali krajowej
Institute for Scientific Information Instytut Informacji Naukowej	do sporządzania indeksu cytatów naukowych
Library of Congress Biblioteka Kongresu	do sporządzania katalogów danych odczytywanych maszynowo
NASA Krajowa Agencja d.s. Przestrzeni Kosmicznej i Astronautyki	do sporządzania NASA — STAR
National Library of Medicine Narodowa Biblioteka Medyczna	do sporządzania Index Medicus i Medlars (system analizowania i wyszukiwania w zakresie literatury medycznej)
Office of Education Urząd dla spraw Oświaty	dla Ośrodka Informacyjnego Badań Oświatowych

Systemy referencyjne (bibliograficzne). Systemy zbiorów danych podstawowych oferują sformalizowane zapisy o określonej strukturze. Opisy wielu materiałów-fotografii, rysunków technicznych, książek itp. — wymagają znacznie bardziej złożonej struktury danych. Systemy wyszukiwania informacji o

tych materiałach można by nazwać systemami referencyjnymi, służą one bowiem do gromadzenia i przechowywania oraz wyszukiwania raczej referencji o danym materiale, a nie jego treści. Na ogół zapisy zawierają w części opisowej terminy przedmiotowe, które przedstawiają treść danego materiału i określają kryteria wyboru. Tego rodzaju dane trudniej jest wyszukiwać aniżeli dane, którymi zajmują się systemy operowania kartotekami, a to z powodu złożonego charakteru powiązań między tematami. System referencyjny musi odznaczać się także zdolnością do operowania tezaurusem, zawierającym strukturę słownika haseł przedmiotowych i wskazującym dla każdego hasła inne wyrazy ogólniejsze i bardziej szczegółowe, podobne lub synonimowe itp. Na ogół objętość kartotek w systemach referencyjnych tak się rozrasta, że w przypadku zapytania, można zbadać zaledwie tylko część zapisów. Opracowuje się dla nich podkartoteki i indeksy (w przypadku krańcowym — jeden dla każdego tematu) tak, aby przy użyciu komputera wydawane były tylko referencje dotyczące zadanego tematu.

Pierwsze referencyjne systemy wyszukiwawcze opracowane zostały dla celów informacji naukowej i technicznej. Indeksy sporządzane z pomocą maszyn matematycznych publikuje szereg ośrodków informacji technicznej, z których największymi są: Ośrodek Dokumentacji Departamentu Obrony wydający TAB, Krajowa Agencja d.s. Astronautyki wydająca STAR oraz Agencja d.s. Energii Atomowej wydająca Abstrakty Atomistyki. W każdym przypadku podstawowy zbiór danych dostępny jest dla uprawnionych użytkowników w celu wyszukiwania referencji. Centralną placówkę gromadzącą informacje naukowo-techniczne w skali federalnej utworzono w Departamencie Handlu USA. Dostarcza ona indeksy porównawcze oraz świadczy usługi w zakresie wyszukiwania referencji.

Z takiej samej techniki, uzupełniającej tradycyjne sposoby katalogowania, korzystają biblioteki. W miarę mechanizacji czynności urzędniczych w bibliotekach — kontroli obiegu, zamawiania książek i czasopism, drukowania katalogów i indeksów itp. zbiory danych podstawowych stały się również wykorzystywane do celów wyszukiwania bibliografii. Narodowa Biblioteka Medyczna publikuje z pomocą maszyny matematycznej *Index Medicus*, miesięczny indeks artykułów w czasopismach medycznych. W systemie bibliograficznym zwanym MEDLARS używa się taśm magnetycznych. Biblioteka Kongresu przystąpiła obecnie do wytwarzania i rozprowadzania zapisów na taśmach magnetycznych zawierających dane katalogowe nabywanych książek. Stowarzyszenia naukowe i zawodowe oraz placówki zajmujące się sporządzaniem abstraktów, w tym *Chemical Abstracts* i *Biological Abstracts* planują wykorzystanie maszyn matematycznych do wydawnictw, zaś w celu wyszukiwania bibliografii będzie można nabywać odpowiednie taśmy. Tablica II podaje reprezentacyjną listę systemów bibliograficznych.

Systemy bibliograficzne są często częścią **systemów wyszukiwania dokumentów.** Zapis wyjściowy systemu bibliograficznego wskazuje, gdzie można znaleźć dany materiał celem udostępnienia go użytkownikowi. W przeszłości większość systemów wyszukiwania dokumentów gromadziła materiały w postaci drukowanej. W ostatnich latach jednak objętość fizyczna wielu zbiorów dokumentów urosła do takich rozmiarów, że zaszła konieczność przechowywania ich w formie mikroobrazowej. Tablica III ilustruje kilka rodzajów takich systemów przechowywania dokumentów.

Wykaz ten bynajmniej nie wyczerpuje wszystkich dostępnych możliwości wyboru. Obejmuje on formy mikrofotograficzne takie jak: taśmy, kasety, karty filmowe (zwane mikrofiszkami), paski a nawet płytki filmowe, jak również kombinacje filmów z innymi formami przechowywania danych. Obejmuje on również przechowywanie mikroobrazów w formie magnetycznej (różnego rodzaju kartoteki wizualne).

Ostatnio poświęca się wiele uwagi przechowywaniu pełnych tekstów dokumentów w postaci magnetycznej dostosowanej do odczytu za pomocą maszyny cyfrowej, gdyż powszechność użycia tych maszyn gwarantuje możliwość udostępniania tekstu w takiej formie.

Systemy przetwarzania tekstów. Przy użyciu maszyn matematycznych do działalności wydawniczej tekst zapisuje się w języku maszyny, zaś za pomocą programów ustala się odpowiednie formaty stron i steruje się przyrządami do fotoskładu, który następnie wydaje gotowe do druku matryce offsetowe.

Z punktu widzenia wyszukiwania informacji, zaleta pełnego zapisu tekstu na taśmie magnetycznej tkwi w możliwościach językowego przetwarzania danych. Niemal klasycznym przykładem jest sporządzanie alfabetycznych spisów wyrazów znaczących wybranych ze wskazaniem kontekstu, w jakim każdy z tych wyrazów występuje. Tego rodzaju techniki używa się w systemach bibliograficznych w celu sporządzania indeksów artykułów na podstawie wyrazów kluczowych w ich tytułach oraz indeksów książek na podstawie ich spisów treści. Obliczenia statystyczne częstotliwości występowania wyrazów w tekście, ich długości, podziałów, powiązań itp., mają duże znaczenie dla językoznawstwa; stanowią pomoc w analizowaniu języka pisanego. Znajdują one również zastosowanie w systemach bibliograficznych. H. P. Luhn wystąpił z poglądem, że względna częstotliwość występowania danych wyrazów w tekście stanowi kryterium, które może służyć maszynie matematycznej do wybierania haseł przedmiotowych określających tekst.

Głębsza analiza tekstu wymaga posługiwania się składnią i semantyką. Próbowano różnych sposobów ułożenia programów maszyny matematycznej celem wyprowadzenia składniowej struktury zdania z syntaktycznych klas wyrazów, z których się to zdanie składa (tak, jak to jest podane w słowniku). Na przykład jeden ze sposobów, zwany **bezpośrednią analizą składnikową** *(immediate constituent analysis)* bierze po prostu pod uwagę tylko kolejność wyrazów w zdaniu. Z tego można następnie wyprowadzić możliwe kolejności klas syntaktycznych tak, aby przynajmniej jedna z tych sekwencji była jedynie możliwym do przyjęcia zdaniem. Inny sposób — to

Tablica III
Mikroobrazowe systemy przechowywania dokumentów

Forma mikroobrazu	Charakterystyka
Rolki filmowe	różna długość, typowa szerokość — 35 mm, zmniejszenie obrazu 20:1
Kasety filmowe	zwykle szerokości 35 mm, długość 100 stóp, zawierają 2000 obrazów stronic w 20-krotnym zmniejszeniu
Mikrokarty	karty, zwykle o rozmiarach 3 × 5 cali, zawierające do 70 matowych obrazów pozytywowych stronic książkowych w 20-krotnym zmniejszeniu
Mikrofiszki	przezroczyste karty, zwykle o rozmiarach 4 × 6 cali, zawierające do 60 i więcej obrazów stronic zmniejszonych 20-krotnie
Karty aperturowe	rodzaj kart perforowanych, z jednym lub wieloma otworami; każda karta zawiera 1 klatkę filmu 35 mm z utrwalonym obrazem dokumentu lub rysunku, typowe zmniejszenie 20 razy
Kartoteki videograficzne obrazowe	taśma magnetyczna z utrwalonymi wizualnymi zapisami stronic dokumentów, zwykle zaopatrzone w dane cyfrowe do identyfikacji

analiza struktury fazowej, która na wzór tradycyjnego rozbioru gramatycznego — traktuje zdanie jako kombinację podstawowych członów zdania, np. „rzeczownikowy trzon zdania" i „czasownikowy trzon zdania". Najbardziej złożonym sposobem jest **teoria transformacji,** traktująca zdania jako odmiany pewnych podstawowych struktur zdaniowych. Przykład: zarówno „Piłka została kopnięta przez Johna", jak i „Czy John kopnął piłkę"? dają się wyraźnie opisać jako stosunkowo proste odmiany zdania „John kopnął piłkę".

Tego rodzaju techniki przetwarzania tekstów są szczególnie przydatne w naukach humanistycznych. Badania nad częstotliwością wyrazów i strukturą zdania stosowane były w analizach stylu, odzwierciedlają one bowiem sposoby używania wyrazów.

Jedną z górnych granic do osiągnięcia w badaniach nad wyszukiwaniem informacji jest wykorzystywanie przetwarzania tekstów i analizy logicznej do znajdywania odpowiedzi na zapytania — poprzez analizę implikacji dających się wyprowadzić ze stwierdzeń zawartych w tekstach. W szeregu eksperymentalnych systemów **pytanie — odpowiedź** dowodzono tego rodzaju możliwości oraz zbadano wynikające z tego techniczne problemy.

Charakterystyczne problemy systemów wyszukiwania informacji

W procesie wdrażania każdego z systemów wyszukiwania informacji wyłania się szereg odmiennych i bardziej złożonych problemów w stosunku do bardziej tradycyjnych zadań z dziedziny ETO lub przetwarzania danych gospodarczych. Z grubsza można by wyróżnić cztery grupy problemów: użyteczność, komunikacja, dobór charakterystyk wyszukiwawczych i organizacja zbiorów.

Czołowy i przy tym najtrudniejszy do rozwiązania — to problem u ż y t e c z n o ś c i tych systemów. Ich rola — to zbieranie i przechowywanie danych wytworzonych dla jakichś określonych celów, danych, mających następnie służyć innym, na ogół trudnym do przewidzenia celom. Tak więc — dane gospodarcze powstające w ramach czynności buchalteryjnych zachowywane są dla późniejszych decyzji z dziedziny zarządzania; artykuły naukowe, wydrukowane dla danego grona czytelników są gromadzone i opracowywane dla użytku innych czytelników itp. O ile jednak początkowe zapisy danych miały na pewno ściśle określoną wartość, związaną z celem ich powstania, to ich użyteczność dla innych, nieznanych celów nie jest całkowicie jasna. Co więcej, koszty zdobywania danych w stadium ich powstawania stanowią na ogół najważniejszy wzgląd. Biorąc pod uwagę fakt, że jest to „nakład kapitałowy", to powinien się on — w pewnym sensie — zamortyzować w toku przyszłego, na razie niewiadomego użytkowania. A zatem — trudno jest uzasadnić wydatki na system wyszukiwania informacji. Co więcej, w jego działaniu nieunikniony jest wysoki stopień niepewności, wynikający nie tylko z przyszłego sposobu użytkowania danych, lecz także ze stopnia ich rzetelności. W samej rzeczy — jeden z głównych problemów polega na tym, aby ustalić, jakie dane zostaną włączone do danego systemu.

Inna grupa zagadnień — k o m u n i k a c j a — wiąże się z faktem, że nagromadzone dane muszą być dobierane tak, aby opisywały fakty zewnętrzne w stosunku do systemu w formie adekwatnej dla tych późniejszych, niepewnych celów. Opisywanie ludzi, sygnałów lub idei nie jest sprawą prostą, nigdy bowiem nie jest jasne, jakie dane będą potrzebne ani też w jaki sposób mają one być ujęte. Podobny problem wyłania się przy nadawaniu zapytaniom formy, w jakiej można by się nimi dalej posługiwać. W przypadku systemów danych podstawowych może to po prostu oznaczać posługiwanie się księgami kodów; w systemach referencyjnych wymaga to na ogół układania słowników i tezaurusów, które — jak to wskazano powyżej — uwidoczniają powiązania pomiędzy hasłami. W systemach przetwarzania tekstów — wewnętrzna logika niezbędna do komunikacji staje się tak dalece skomplikowana, że wielu badaczy utrzymuje, iż wykracza ona poza ramy obecnej wiedzy o strukturze języka oraz możliwości istniejących komputerów.

Trzecia grupa problemów — d o b ó r c h a r a k t e r y s t y k w y s z u k i w a w c z y c h — wynika z faktu, że celem wyszukiwania jest wybór właściwych (*„relevant"*) danych ze zbioru. Pojęcie relewancji jest jednakże tak niejednoznacznie zdefiniowane, że na ogół nie jest jasne, jak należy ją mierzyć. Co więcej, niepewność, a nawet błędy w opisie zarówno przechowywanych danych, jak i w zapytaniach są tak znaczne, że właściwie nie jest jasne, czy ścisłe miary relewancji — zgodności zapytania z odpowiedzią — mają jakieś realne znaczenie. Ogólnie biorąc, celowe jest umożliwienie kolejnego doboru charakterystyk wyszukiwawczych, przy czym w ostatecznym rachunku decyduje użytkownik. Każdy z tych doborów służy jako sito ograniczające ilość zapisów do zbadania w toku dalszych przesiewów. Przy takim postępowaniu wzrasta prawdopodobieństwo przepuszczenia poszukiwanych tematów, które mogą być odrzucone przez jedno sito, nawet wtedy, gdyby zostały zaakceptowane w toku dalszego, bardziej uważnego przesiewu. I znów — ponownie dużej wagi nabierają trudności, wynikające z jakości pierwotnych opisów; nie wszystkie bowiem dane opisujące mogą być lub będą włączone, a ważne związki między hasłami opisującymi mogą być pominięte. Problemy komunikacji i problemy doboru charakterystyk wyszukiwawczych razem wzięte wywołały u wielu ludzi zajmujących się systemami wyszukiwania danych przekonanie, że wzajemne oddziaływanie człowiek — maszyna staje się niezbędne. W rezultacie należało wiele uwagi poświęcić wykorzystaniu techniki obliczeniowej *„on line"*.

Czwarta grupa problemów — o r g a n i z a c j a z b i o r ó w — wypływa z faktu, że w przypadku każdego ważniejszego zastosowania kartoteki rozrastają się do takich rozmiarów, iż staje się niemożliwe, a przynajmniej ekonomicznie niecelowe, szczegółowe uwzględnianie w nich każdego tematu. Dla przykładu warto wskazać, że — o ile wielka macierz taka, jaką otrzymuje się przy rozwiązywaniu równania różniczkowego o pochodnych cząstkowych, wymagałaby najwyżej 10^7 lub 10^8 znaków przechowawczych, a bardzo obszerna ewidencja inwentarzowa 10^8 lub 10^9 znaków przechowawczych, to katalog średniej nawet biblioteki uniwersyteckiej wymagałby 10^9—10^{10} znaków. Dla obszerniejszych zagadnień rozmiary kartoteki do wyszukiwania informacji mogą sięgać aż 10^{11}—10^{12} znaków.

W przypadku kartotek o tak dużych rozmiarach trzeba by stworzyć dla nich specjalną strukturę, przechowywać je na różnych szczeblach pamięci, opracować dla nich mechanizm indeksowy oraz stosować kolejne przesiewy umożliwiające zredukowanie ilości zapisów, które mają podlegać szczegółowym badaniom. To właśnie ta czwarta grupa problemów stanowi w ostateczności rzeczywisty problem techniczny. Tu bowiem oddziaływają na siebie wzajemnie wszystkie sprawy związane z rozmiarami kartotek, niezbędnym czasem udzielania odpowiedzi, dostępnym sprzętem, stopniem selektywności oraz ze ścisłością odpowiedzi.

Podsumowanie

Reasumując: wyszukiwanie informacji stało się terenem szerokiego zastosowania komputerów. Wprawdzie możliwe jest wyodrębnienie różnorodnych systemów, to jednak wszystkie one — w mniejszym lub większym stopniu — związane są z podstawowymi problemami: niepewne użytkowanie niepewnych danych, ogrom rozmiarów kartotek, złożone struktury danych oraz nie oznaczona użyteczność ekonomiczna. W miarę jednak stosowania — problemy te są rozwiązywane i udało się już stworzyć pożytecznie działające systemy.

Andrzej MAKOWSKI
Stołeczny Ośrodek Elektronicznej Techniki Obliczeniowej
Warszawa

681.322.004.016:002.513.5

INBI

System przetwarzania na EMC informacji bibliograficznych

Pierwszy polski zautomatyzowany system informacji bibliograficznej INBI, z zastosowaniem EMC ZAM-41, jest od 1967 r. wykorzystywany do redagowania miesięcznika „Automatyzacja Przetwarzania Informacji — Bibliografia", wydawanego przez Ośrodek Informacji Naukowo-Technicznej i Ekonomicznej Instytutu Maszyn Matematycznych w Warszawie. W artykule opisano: organizację systemu INBI opartą na klasyfikacji hierarchiczno-przedmiotowej, sposób wprowadzania danych (opisów bibliograficznych) na papierowej 8-kanałowej taśmie perforowanej za pośrednictwem automatu organizacyjnego typu „Soemtron", ograniczenia systemu i postać wyników.

Wprowadzenie

System INBI — akronim od INformacja BIbliograficzna — został opracowany [1] w celu automatycznego przetwarzania informacji bibliograficznych włączanych do miesięcznika „Biuletyn Informacyjny IMM — Bibliografia", redagowanego przez Sekcję Dokumentacji Działu Informacji Naukowej i Wydawnictw Instytutu Maszyn Matematycznych [2].

Według pierwotnego zamiaru, system INBI miał realizować następujące funkcje:

1) redakcja miesięcznika
2) redakcja indeksu rocznego
3) wyszukiwanie informacji
4) sporządzanie statystyk

Przed rozpoczęciem prac nad systemem, z wymienionych wyżej funkcji realizowana była tylko jedna — redakcja miesięcznika. Pozostałe miały stanowić w pewnym sensie *novum* w stosunku do dotychczasowej pracy Sekcji Dokumentacji.

W pierwszej połowie 1967 roku wdrożono system INBI/A na EMC ZAM-21, z programami napisanymi w Podstawowym Języku Programowania, z informacjami bibliograficznymi perforowanymi na 5-kanałowej taśmie papierowej. System ten został przygotowany do realizacji trzech pierwszych z wymienionych funkcji. W następnej kolejności wdrożono system INBI/B na EMC ZAM-41 (przygotowano program redagujący miesięcznik) z programami pisanymi w Podstawowym Języku Symbolicznym, z informacjami dziurkowanymi na 8-kanałowej taśmie, z wykorzystaniem taśm magnetycznych. Różnica między obu wersjami jest najbardziej widoczna w samym wydawnictwie: system INBI/A pozwalał na użycie w tekstach tylko wielkich liter (zeszyty od nr 5/67 do nr 7/68); system INBI/B dopuszcza zarówno wielkie, jak i małe litery (zeszyty od nr 8—12/68).

Zarys dotychczasowego systemu organizacji informacji bibliograficznej [3]

Do redagowania miesięcznika „Bibliografia" służyły **„opisy bibliograficzne"** ważniejszych publikacji dotyczących tematyki elektronicznych maszyn cyfrowych.

Opisy bibliograficzne **indeksowano** za pomocą tablic klasyfikacyjnych, sporządzonych przez Sekcję Dokumentacji.

Tablice te, składające się z działów i poddziałów, specjalnie dostosowanych do zagadnień maszyn matematycznych stanowią twór samoistny, nie oparty na Uniwersalnej Klasyfikacji Dziesiętnej; są one w toku prac Sekcji poddawane stałej weryfikacji.

Opisy przepisywano na karty dokumentacyjne, zaś po miesięcznym okresie **sortowano** je (ok. 250 sztuk) według działów i poddziałów oraz zaopatrywano (numeratorem) w kolejny numer pozycji bibliograficznej. Ciągłość numeracji jest utrzymywana od początku wydawnictwa, tj. od września 1961 r.

W czasie przygotowywania miesięcznika wypisywano kartki **„odsyłaczy"** zawierające skrócony opis oraz numer pozycji opisu, do którego odsyłacz się odnosi i umieszczano je w odpowiednich działach (poddziałach).

Tak przygotowany zbiór kart z opisami przepisywano po kolei na matryce w celu powielenia, wypisując jednocześnie pełne brzmienie działów i poddziałów wraz z ich symbolami.

Po sporządzeniu kolejnego zeszytu miesięcznika, karty włączane były do Kartoteki-matki jako materiał dokumentacyjny.

Cele systemu INBI

Opracowanie systemu INBI miało na celu postawienie pierwszych kroków w zastosowaniu EMC do przetwarzania informacji bibliograficznej w Polsce. Ze względu na nowość tego rodzaju prac i brak doświadczeń postanowiono przy pomocy systemu INBI uzyskać wyniki (w zakresie czynności redagowania miesięcznika „Bibliografia") możliwie najbardziej zbliżone do dawnej formy wydawniczej. W ten sposób wielokrotnie zwiększono szanse powodzenia prac nad systemem i to zarówno od strony dotychczasowej praktyki, jak i od strony psychologicznej; zredukowano do minimum trudności adaptacyjne człowieka mającego współpracować z EMC (co nawet przy tak stosunkowo błahym problemie dawało się zauważyć).

Wymagania wobec EMC

System INBI/A został opracowany na EMC ZAM-21. Egzemplarz maszyny, którym wówczas dysponował Instytut Maszyn Matematycznych, był wyposażony w pamięć operacyjną o pojemności 8 K słów 24-bitowych oraz w pamięć bębnową o pojemności 32 K słów 24-bitowych. Chcąc opracować maszynowo jeden numer „Bibliografii" trzeba było oszacować, z ilu znaków pisarskich numer się składa: 250 opisów × 8 wierszy × 40 znaków w wierszu = 80 000 znaków. Ponieważ w 24-bitowym słowie maszyny mieszczą się 4 znaki pisarskie kodowane na 5-kanałowej taśmie, a więc na 80 000 znaków potrzeba 20 K słów maszyny, co wskazuje na możliwość umieszczenia na bębnie całego tekstu „Bibliografii" (pamięć operacyjna w tym przypadku służy tylko do opracowania pojedynczych opisów). Jednocześnie było widać, że nie można nadmiernie zwiększyć ilości opisów albo ich długości, gdyż wtedy omawiany egzemplarz EMC okazałby się za mały.

---

[1] Koncepcję i założenia opracowali Władysław Klepacz i Dorota Prawdzic. Założenia programowe, programy napisał oraz system INBI wdrożył Andrzej Makowski, przy współudziale Sekcji Dokumentacji Działowego Ośrodka Informacji NTE IMM.

[2] Od stycznia 1968 r. miesięcznik ukazuje się pod nazwą „Automatyzacja Przetwarzania Informacji — Bibliografia", wyd. Działowy Ośrodek INTE IMM.

[3] Patrz także artykuł: D. Prawdzic i A. Senkowski — „Działalność Ośrodka Informacji Naukowo-Technicznej elektronicznej techniki obliczeniowej", „Maszyny Matematyczne", nr 1/65, str. 24.

Sytuację radykalnie zmieniło użycie taśm magnetycznych w systemie INBI/B na ZAM-41: na jednym krążku taśmy mieści się kilka milionów słów 24-bitowych.

Z urządzeń wejścia-wyjścia maszyny w systemie INBI/A wykorzystane były:

1) czytnik 5-kanałowej taśmy perforowanej
2) 5-kanałowy perforator taśmy papierowej
3) drukarka wierszowa (odgrywająca zasadniczą rolę w szybkim wydruku wyszukanych informacji)

zaś w systemie INBI/B:

1) czytnik 5-kanałowej taśmy perforowanej
2) czytnik 8-kanałowej taśmy perforowanej
3) 5-kanałowy perforator taśmy papierowej
4) 8-kanałowy perforator taśmy papierowej
5) drukarka wierszowa
6) 3 (lepiej 4) stacje taśm magnetycznych.

Wymagania wobec systemu organizacji informacji bibliograficznej

Podstawowym wymaganiem przy projektowaniu wszelkiego rodzaju systemów elektronicznego przetwarzania danych, stawianym użytkownikom — jest podanie definicji lub umożliwienie **jednoznacznego zdefiniowania** dotychczasowego **systemu.**

W przypadku stosowanego dotychczas systemu informacji bibliograficznej daje się zauważyć szczególną trudność, mianowicie brak zdefiniowania — na dłuższy okres czasu — działów i poddziałów klasyfikacji. Zmiany w klasyfikacji uniemożliwiają sporządzenie prawidłowych indeksów (przynajmniej za okres, w którym zaszły jakiekolwiek zmiany w klasyfikacji), czy też wyszukiwanie informacji stosowane przy większych zbiorach, czyli obejmujące dłuższe okresy. Istnieją jednak **dopuszczalne zmiany w klasyfikacji**, mianowicie takie, które niczego w dotychczasowych tablicach nie zmieniając, stanowią ich rozszerzenie. A więc możliwe jest wyodrębnianie nowych działów i poddziałów.

Niezawodność systemu przetwarzania

Maszynowe opracowanie wydawnictwa „Bibliografia" pozwoliło na wyeliminowanie błędów, które w sposób nieunikniony wkradały się do numerów, opracowywanych ręcznie.

Np. w czterech pierwszych numerach „Bibliografii" w r. 1967 — nr 5 był już opracowany maszynowo — zauważono następujące błędy:
nr 1, str. 2: przestawiono numery bibliograficzne odsyłaczy — 12878, 12877
str. 5 — opuszczono symbol poddziału: 71 (teoria kodów)
str. 8 — przestawione numery bibl. odsyłaczy: 12878, 12877
str. 14 — niewłaściwy nr bibl. 12924 zamiast 12921
nr 2, str. 4 — niewłaściwy nr bibl.: 12969 zamiast 12963
str. 6 — przesunięcie pierwszego wiersza pozycji 13211 o jedną spację w lewo
str. 18 — przestawione nr bibl. odsyłaczy: 12975, 12095
str. 22 — niewłaściwy symbol poddziału: 211 zamiast 29

nr 3, str. 20 — brak pozycji o nr bibl. 13325
str. 21 — brak pozycji o nr bibl. 13339
str. 24 — dwukrotnie wpisany nr bibl. 13361 wraz z częścią opisu
str. 25 — brak pozycji o nr bibl. 13373
— brak pozycji o nr bibl. 13374
nr 4, str. 14 — brak pozycji o nr bibl. 13483.

Chodzi tu oczywiście **o błędy typu formalnego;** błędy rzeczowe, popełnione przez opracowującego opisy bibliograficzne (wraz z indeksowaniem) nie ulegają eliminacji, zarówno w redakcji ręcznej, jak i maszynowej.

Opracowanie maszynowe nie usuwa całkowicie błędów formalnych, prawdopodobieństwo jednak występowania tego rodzaju błędów jest o wiele mniejsze. Należałoby tu zaznaczyć, że możliwość „popełniania błędu" przez część centralną (elektroniczną) EMC jest zupełnie znikoma; jeżeli w opracowaniu maszynowym zdarzają się błędy, to powstają one w urządzeniach zewnętrznych maszyny — (mechanicznych) czyli w urządzeniach wejścia-wyjścia, najczęściej na skutek przekłamania odczytu taśmy perforowanej w czytniku lub w dalekopisie.

W pierwszych czterech numerach „Bibliografii", opracowawych automatycznie, spostrzeżono następujące błędy:
nr 5 —, nr 6, str. 14 — opuszczony opis poz. bibl. o nr poz. 14094 — zamiast niego wypisane cyfry 4 oraz 5
nr 7, str. 1 (tytułowa, redagowana ręcznie) — niewłaściwy numer ostatniej pozycji Biuletynu 14385 zamiast 14387
str. 6 — przesunięcie pierwszego wiersza poz. 14194 o jedną spację w prawo, nr 8.

Błędy w nr 6 „Bibliografii" powstały na skutek przekłamania czytnika maszyny: pominięty został początkowy znak separatora, stojący przed opisem bibliograficznym, zaś końcowy separator (po opisie) został potraktowany jako początkowy, w wyniku czego miejsce właściwego opisu bibliograficznego zajęły cyfry, dotyczące odsyłacza do omawianej pozycji. Wyżej omówione przekłamanie spowodowało także brak odsyłacza w dziale IV, poddziale 5. Pomyłki tej można było uniknąć przez dokładniejsze wizualne sprawdzenie Biuletynu i powtórne opracowanie go przez EMC.

Łatwo zauważalnym błędem graficznym powstałym w wyniku dodawania spacji na początku niektórych wierszy przez dalekopis było „rozmycie" kolumn w indeksie do „Bibliografii" za rok 1967. Można było temu zapobiec przez wypisanie indeksu na innym dalekopisie.

Możliwość różnorodnych opracowań

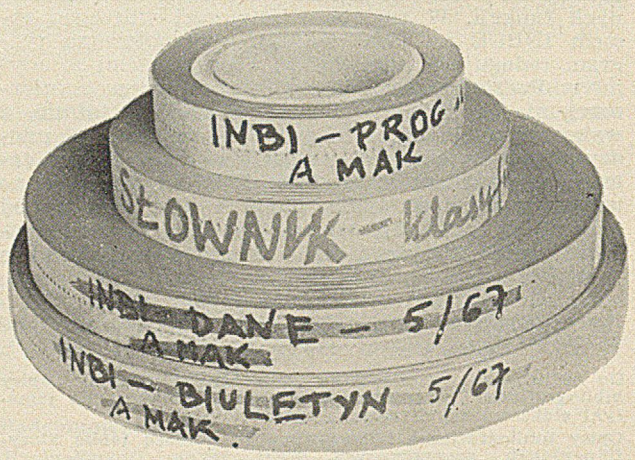
Po przygotowaniu opisów bibliograficznych w formie danych wejściowych nadających się do wprowadzenia do EMC, mamy możliwość różnorodnego i wielokrotnego ich wykorzystania przez różne programy systemu. I tak w systemie INBI/A raz przygotowane dane wejściowe były wykorzystywane do:

1) redagowania miesięcznika
2) sporządzania indeksu.

Planowane wykorzystanie omawianych danych do wyszukiwania informacji i sporządzania statystyk nie zostało zrealizowane ze względu na zbyt jeszcze krótki okres użytkowania (za mały zbiór). Dane stają się podstawą różnorodnych opracowań po przygotowaniu odpowiednich programów.

Wybór nośnika informacji bibliograficznych

W problemach przetwarzania danych nośnikiem informacji wejściowej jest powszechnie używana papierowa taśma perforowana 5-ścieżkowa, względnie karty dziurkowane 80- albo 90-kolumnowe.



Rys. 1. Krążek taśmy systemu INBI

Jak już było powiedziane wyżej, dane do jednego numeru „Bibliografii" składają się z ok. 80 000 znaków pisarskich, co przy gęstości zapisu informacji **na taśmie** równej 4 znaki na cm bieżący, oznacza ok. 200 m taśmy czyli krążek o średnicy ok. 20 cm (rys. 1). Ta sama ilość informacji wydziurkowana **na kartach** 80-kolumnowych zajęłaby aż 1000 kart.

I ZAGADNIENIA OGÓLNE M. L.

13388. Die Generationsfolge in der elektronischen Datenverarbeitung.
SCHNEIDER C.
Cechy charakterystyczne trzech generacji EMC: urządzenia EMC i systemy programowania.
BTO, 1967.02:15/2/, s. 114, 116-119. Sygn. 0154.

2 Ośrodki obliczeniowe i badawcze

Zob. też poz. 13633.

31 Europa

311 Anglia

13392. Some recent UK orders.
Zamówienie brytyjskie na EMC serii GE-400 i GE-115.
J Bull Syst, 1967:/19/, s. 10-11. Sygn. 0613.

Rys. 2. Fragment strony Biuletynu Informacyjnego „BIBLIOGRAFIA" (metoda ręczna)

1 ZAGADNIENIA OGOLNE M. L.

1 ZAGADNIENIA HISTORYCZNE I PRZYSZLOSCIOWE

13638. VYCISLITEL'NAJA TECHNIKA I MECHANIZACIJA UPRAVLENCESKOGO TRUDA. SB. 2.
1966, LENINGRAD, UNIV., LENINGRAD, S. 109.
S. 75: NEKOTORYE NAPRAVLENIJA RAZVITIJA ELEKTRONNYCH CIFROVYCH MASIN.
GORAL'NIK L.A.
S. 75-87, POZ. BIBL. 9. SYGN. B 426/2.

3 PANSTWA. FIRMY. ORGANIZACJE

31 EUROPA

312 FRANCJA

13646. FRENCH PRESS BAITS BULL.
KILBORN P.
NOWE TRENDNOSCI PRODUKCYJNE FIRMY BULL.
ELECTRONICS, 1967.04:40/7/, S. 163-164, 166-167.
SYGN. 0500.

318 KRAJE SOCJALISTYCZNE

ZOB. POZ. 13757

Rys. 3. Fragment strony Biuletynu Informacyjnego „BIBLIOGRAFIA", wykonanego systemem INBI/A

VIII.3 Projektowanie systemów mechanizacji i automatyzacji zarządzania

18282. Efektywność przetwarzania danych gospodarczych.
SOWA K.
Efektywność różnych technik przetwarzania danych gospodarczych z punktu widzenia rachunkowości przedsiębiorstwa przemysłowego.
1968, PWE, W-wa, ss. 275. Sygn. 9739.

18283. System informacji statystycznej dla handlu hurtowego przy zastosowaniu EPD.
JE-ROZYŃSKA H.
System informacji /ewidencji i sprawozdawczości/ dla zarządzania handlem hurtowym z zastosowaniem EPD. Ogólne założenia Prace przygotowawcze.
1968, IHZ, W-wa, ss. 111. Sygn. 9716.

Zob. też poz. 18053

IX AUTOMATYZACJA DOKUMENTACJI I PRZEKŁAD MASZYNOWY

IX.1 Automatyzacja i mechanizacja prac dokumentacyjnych i bibliotekarskich

18288. IFIP Congress 68 Edinburgh - August 68. Invited papers.
1968, North-Hol-land, Amsterdam, ss.270.
S. 221: Search and retrieval experiments in real-time information retrieval.
SALTON G.
Wyszukiwanie informacji w czasie rzeczywistym /systemy Medlars, Smart/.
S. 221-232, poz. bibl. 15. Sygn. 9700d.

Rys. 4. Fragment strony miesięcznika „API-BIBLIOGRAFIA", wykonanego systemem INBI/B

W systemie INBI/A 5-kanałowa taśma perforowana na dalekopisach typu Lorenz wprowadza pewne **ograniczenia w używanych literach.** Zezwala na zakodowanie tylko 60 różnych znaków pisarskich, co wyklucza zakodowanie małych liter oraz liter ze znakami diakrytycznymi.

System INBI/B daje o wiele większe bogactwo znaków, dzięki automatom organizacyjnym typu „Soemtron" lub „Optima" perforującym **8-kanałową taśmę,** na której można zakodować aż 512 znaków. W tej sytuacji, oczywiście, nie wszystkie możliwości taśmy są wykorzystane.

Informacje podstawowe o budowie systemu INBI

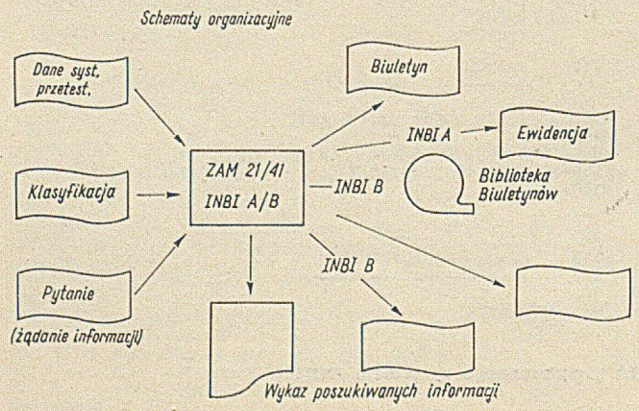
● **INBI/A**

1. Elektroniczna maszyna cyfrowa ZAM-21
2. Język systemu: Podstawowy Język Programowania PJP
3. Stosowany kod: Międzynarodowy drugi M-2

● **INBI/B**

1. EMC ZAM-41
2. Język systemu: Podstawowy Język Symboliczny PJES
3. Kod programów oraz pytań przy wyszukiwaniu informacji: M-2
4. Kod klasyfikacji oraz danych: BCD Optima, stosowany w automatach organizacyjnych z pewnymi ograniczeniami.

Schematy organizacyjne systemu INBI przedstawia rys. 5.



Rys. 5. Schematy organizacyjne

Dane wejściowe, ich źródła i dostosowanie do systemów INBI

Podstawą danych wejściowych systemu są karty dokumentacyjne, zawierające opisy bibliograficzne.

Przykładową kartę z opisem bibliograficznym przedstawia rys. 6. Informacje na karcie stanowi:

1. Nagłówek główny opisu

a) symbol liczbowy (cyfry rzymskie) działu, do którego dany opis bibliograficzny się odnosi

b) symbol liczbowy (cyfry arabskie) poddziału, do którego dany opis bibliograficzny się odnosi

2. Nagłówki odsyłaczy (jeżeli dany opis bibliograficzny odnosi się do kilku działów — poddziałów) o budowie identycznej, jak budowa nagłówka głównego

3. Tekst opisu bibliograficznego.

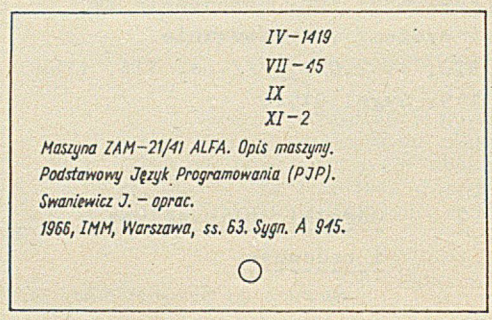
Aby opis bibliograficzny stał się użyteczny do celów wyszukiwania według założonych kryteriów zaopatrzono go w dodatkowe trzy (w INBI/B — cztery) informacje, wypisywane obok nagłówka głównego:

1) dwie ostatnie cyfry roku wydania publikacji

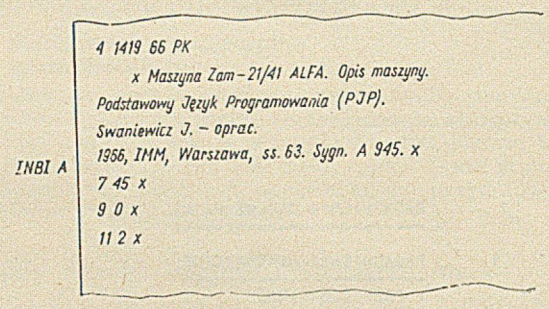
2) symbol języka publikacji

3) symbol rodzaju dokumentu zawierającego publikację

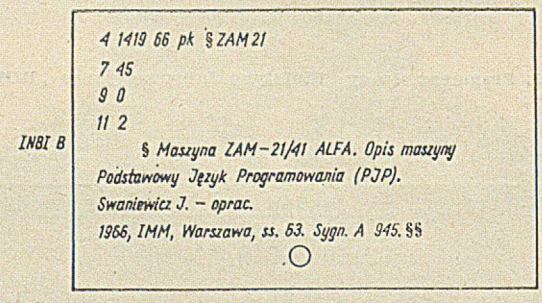
4. — w INBI/B — kryterium alfanumeryczne (trzecie kryterium sortowania[4]) zbioru opisów bibliograficznych — po dziale i po poddziale) — jeżeli ma ono przy danym opisie bibliograficznym zastosowanie.



Rys. 6. Przykładowa karta z opisem bibliograficznym



Rys. 7. INBI/A



Rys. 8. INBI/B

Kryterium alfanumeryczne jest:

kryterium pozycyjnym (pozycje adiustowane do lewej strony)

kryterium, w którym litera **Z** wyprzedza na danej pozycji dowolną cyfrę (np. 301 jest wcześniejsze od 7, CDK21 jest wcześniejsze od CD207).

---

4) Sortowaniem nazywamy porządkowanie dowolnych zbiorów, dające się sprowadzić do porządkowania liczb relacją mniejszości (tzn. po uporządkowaniu każda liczba ma być większa od poprzedniej, zaś mniejsza od następnej — względnie odwrotnie).

Symbole języków publikacji	Symbole rodzajów dokumentów
p — polski	k — książki
a — angielski	c — czasopisma
r — rosyjski	m — mikrofilmy
n — niemiecki	p — prospekty
f — francuski	i — inne
c — czeski	
i — inne języki	

Wprowadzanie danych wejściowych

W systemie INBI/A należy informacje z karty dokumentacyjnej przenieść na taśmę perforowaną, ujmując tekst opisu w znaki separatora (znak gwiazdki *).

W systemie INBI/B już same karty opisów bibliograficznych można wypisywać na automatach organizacyjnych typu „Soemtron" lub „Optima", perforując jednocześnie w kodzie BCD 8-kanałową taśmę papierową, będącą właściwym nośnikiem danych wejściowych systemu. Kryterium alfanumeryczne należy podać po znaku separatora (znak paragrafu §). Tekst opisu należy ująć w znaki separatora: przed tekstem jeden znak, po tekście — dwa.

Dla obu systemów liczbę reprezentującą dział należy zapisać cyframi arabskimi.

Na początku taśmy z danymi trzeba umieścić **nagłówek taśmy** złożony z:

1) numeru kolejnego pierwszej pozycji zeszytu „Bibliografii"

2) numeru miesięcznika

3) dwóch ostatnich cyfr roku wydania danego zeszytu.

Na końcu ostatniego opisu bibliograficznego, przeznaczonego do przetwarzania należy wypisać dodatkowy znak separatora (rys. 7 i 8).

Jaką formę nadamy tekstowi na taśmie z danymi, taką będzie miał tekst w wydawnictwie.

Ograniczenia

Informacje wyczytywane do maszyny są umieszczane w słowach lub w częściach słów o ilości bitów odpowiadającej danej informacji.

W systemie INBI/B przyjęto następujące ograniczenia:

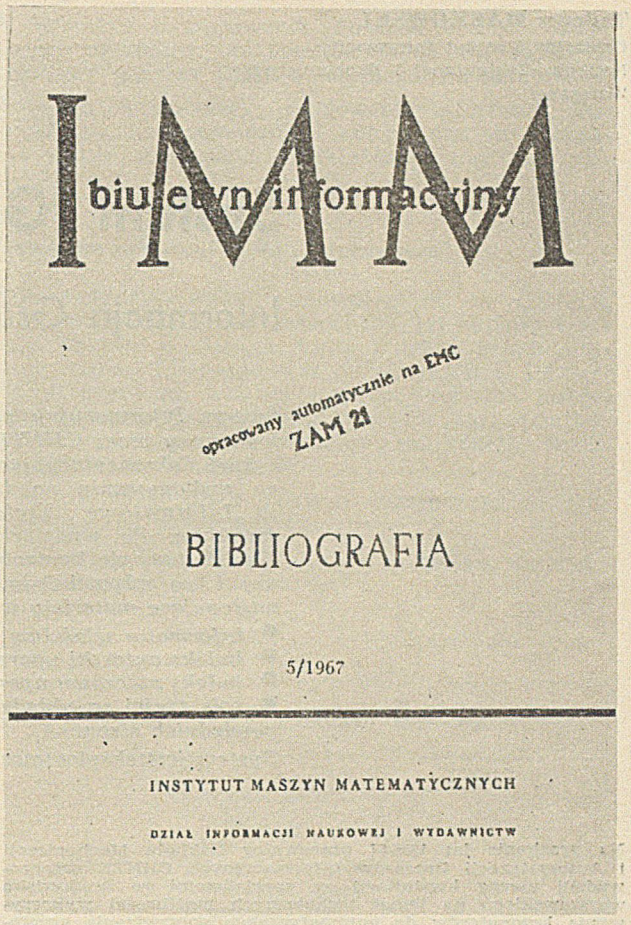
Rok wydania danego zeszytu „Bibliografii"	— 66 < liczba < 82
Nr zeszytu	— liczba < 16
Nr kolejny pozycji bibliograficznej	— liczba < 1 048 576
Dział	— liczba < 16
Poddział	— liczba < 1 048 576
Rok publikacji	— liczba < 128

Kryterium alfanumeryczne — ciąg najwyżej 8 znaków alfanumerycznych (litery duże prócz Ł i Ż oraz cyfry — najpierw litery)

Tekst opisu bibliograficznego — ciąg znaków znaczących	< 1600
ilość nowych linii w tekście opisu	< 32

Wyniki

Wyniki z maszyny wypisywane są w systemie INBI/A i B na takich nośnikach, na jakich były wprowadzone dane. Wyjątek stanowi tu wyszukiwanie informacji, w którym zależy na szybkim uzyskaniu czytelnej (nie zakodowanej) odpowiedzi, co można uzyskać na drukarce wierszowej. W systemie INBI/B można otrzymywać odpowiedzi zarówno na drukarce, jak i na perforatorze — z późniejszym odczytaniem taśmy i wypisaniem tekstu na automacie organizacyjnym. Pierwsza forma jest uboższa w znaki pisarskie (odpowiada taśmie 5-kanałowej).
Druga forma — obok zastosowania jako matryca do powielania — mogłaby być stosowana przy pisemnych odpowiedziach na kwerendy nadsyłane do Działu Informacji.
Wykonane za pomocą systemu INBI biuletyny miesięczne oraz indeksy roczne są periodykami rozsyłanymi do prawie wszystkich instytucji w Polsce, zajmujących się problemami elektronicznej techniki obliczeniowej.



Rys. 9. Okładka pierwszego zeszytu „BIBLIOGRAFII", wykonanego systemem INBI/A

Natomiast wykazy poszukiwanych informacji miały być sporządzane w miarę potrzeby, na konkretne życzenie zainteresowanego. Ze względu jednak na szczupłość opracowanego na EMC zbioru pozycji bibliograficznych w systemie INBI/A (tylko rok 1967) oraz na perspektywę znacznie ulepszonego wyszukiwania informacji w systemie INBI/B, program wyszukujący w systemie INBI/A miał charakter programu wyłącznie pokazowego.

Pytanie-żądanie informacji musiało zawierać sprecyzowany uprzednio temat w postaci liczb działu i poddziału oraz ewentualnie przedział lat, w którym szukane wydawnictwa mogły się ukazać, a także symbole zadanych języków i rodzajów publikacji. Wykaz poszukiwanych informacji otrzymywało się na drukarce wierszowej w postaci zestawu trójek liczb (jak w indeksie), uporządkowanych według lat, języków i rodzajów wydawnictw.

*
* *

Wydaje się, że obok prac nad wprowadzaniem doskonalszych systemów indeksowych typu KWOC (KWIC) czy też systemów opartych na tezaurusach — słuszna jest akceptacja maszynowych opracowań, szeroko stosowanych systemów informacyjnych opartych na klasyfikacji.

I tak np. stwierdzono możliwość zastosowania systemu INBI (po odpowiedniej modyfikacji) do opracowania „Informatora o programach dla maszyn matematycznych", wydawanego przez CO PAN — (ciekawe statystyki o wykorzystaniu różnych EMC) oraz „Przeglądu Dokumentacyjnego Geologii" wydawanego przez Branżowy Ośrodek INTE Instytutu Geologicznego w formie wydawniczej niemal identycznej z formą biuletynu „API — Bibliografia".

Tadeusz MARKOWSKI
Centralny Instytut Informacji
Naukowo-Technicznej i Ekonomicznej
Warszawa

681.322.004.14:002.513.5:058

System IGA

Informacja Grupowana Automatycznie

System informacyjny pod nazwą „IGA" (Informacja Grupowana Automatycznie) wykonany przez CIINTE i ZOWAR polega na sporządzaniu na maszynie IBM 1440 z kart dokumentacyjnych „Informatora o zakończonych pracach naukowych" oraz na wykonywaniu na żądanie zestawień tematycznych wg kryteriów zawartych w „Informatorze". Treść kart dokumentacyjnych jest przeniesiona na karty perforowane. Po wydziurkowaniu partii kart dokumentacyjnych wykonuje się na EMC listowanie kontrolne i nanoszenie poprawek. Po wykonaniu tej pracy wstępnej i po odpowiednim przetworzeniu danych wejściowych system ten tworzy następujące materiały wynikowe:

- *informator właściwy w układzie UKD;*
- *indeks autorski uporządkowany alfabetycznie;*
- *indeks rzeczowy uporządkowany alfabetycznie;*
- *spis treści w układzie UKD z wymienieniem działów i poddziałów oraz odpowiednich stron.*

System jest eksploatowany od początku 1968 roku.

Na przełomie lat 1966/67 pracownicy Zakładu Mechanizacji i Automatyzacji Procesów Informacyjnych CIINTE przeprowadzili szereg konsultacji ze specjalistami ze środowiska warszawskiego na temat technicznych możliwości wykorzystania komputerów do automatycznego sporządzania egzemplarza wydawniczego „Informatora o zakończonych pracach naukowych i naukowo-badawczych" wg UKD (Symboli Uniwersalnej Klasyfikacji Dziesiętnej) oraz do wyszukiwania informacji. Wyszukiwanie informacji miało uwzględniać przy tym takie kryteria, jak symbole UKD, autor, wyraz kluczowy (hasło przedmiotowe).

W wyniku tych konsultacji, na zlecenie Centralnego Instytutu Informacji Naukowo-Technicznej i Ekonomicznej (CIINTE), Zakład Obliczeniowy — Warszawa (ul. Śniadeckich 8) opracował wstępne założenia dla elektronicznego systemu informacyjnego, zwanego wówczas w skrócie „Informator" (obecna nazwa Informacja Grupowana Automatycznie „IGA"). Uwzględniając te założenia, CIINTE powołał zespół roboczy do opracowania projektu systemu. Niezależnie od tego została zawarta odpowiednia umowa pomiędzy CIINTE a Zakładem Obliczeniowym ZOWAR-Warszawa na opracowanie właściwych programów i dokumentacji eksploatacyjnej [1]).

Wykonanie szczegółowego projektu technicznego systemu było, oczywiście, konieczne. Projekt ten objął:

- opracowanie pod kątem technicznym przetwarzania zasadniczych elementów karty dokumentacyjnej, sposobu przenoszenia informacji z karty dokumentacyjnej na karty perforowane, opracowania wzorów kart perforowanych, instrukcji dziurkowania kart;
- opracowanie metody pozwalającej na wykorzystanie przez EMC indeksu UKD mimo jego niejednolitej formy;
- opracowanie postaci wyników tzw. wzorów różnego typu stron informatora: stron „Informatora" właściwego, stron indeksu autorskiego, stron indeksu przedmiotowego, stron spisu treści;
- opracowanie szczegółowego schematu obliczeń z podziałem na poszczególne programy, ustalenie dla każdego z programów:
  - jego funkcji,
  - danych wejściowych,
  - wyników;
- opracowanie metodyki przygotowania organizacyjnego systemu ustalenia terminów dostarczania i tempa spływu danych do dziurkowania, terminów i metody nanoszenia poprawek. W zakres tych prac wchodziłoby również krótkie przeszkolenie osób użytkownika;
- opracowanie kwartalnego terminarza prac;
- podanie wiążących kosztów;
- zaprogramowanie systemu;
- uruchomienie systemu oraz szacowanie kosztu eksploatacji systemu, jak też pracochłonności w.w. prac;
- uzasadnienie efektów ekonomicznych płynących z przejścia na automatyczne przetwarzanie informacji.

Projekt ten stanowił podstawę do zaprogramowania systemu i jego uruchomienie.

Cechy szczególne zbioru informacyjnego i sposoby przygotowywania informacji bieżącej (bibliograficzno-dokumentacyjnej) o tym zbiorze

Rejestr prac naukowych i naukowo-badawczych został powołany do życia na mocy zarządzenia nr 42 Przewodniczącego Komitetu Nauki i Techniki z dnia 10 lipca 1964 r. w sprawie rejestracji zakończenia prac naukowych i naukowo-badawczych. Prowadzenie rejestru zostało powierzone CIINTE.

W celu przypomnienia podaje się, że obowiązkowi rejestracji w CIINTE podlegają:

1) ważniejsze prace naukowo-badawcze w dziedzinie przygotowania nowej techniki oraz ważniejsze prace podstawowe, teoretyczne i doświadczalne w dziedzinie nauk matematyczno-przyrodniczych i społecznych, zakończone po dniu 1 stycznia 1964 r. w jednostkach organizacyjnych podległych Ministerstwu Szkolnictwa Wyższego oraz w placówkach Polskiej Akademii Nauk,

2) wszystkie prace naukowo-badawcze zakończone po dniu 1 stycznia 1964 r. w innych państwowych jednostkach organizacyjnych.

Rejestr prac naukowych i naukowo-badawczych prowadzony jest chronologicznie, tzn. jest uzupełniany w miarę napływu kart dokumentacyjnych do CIINTE, w ilości średnio 500—600 miesięcznie. Zarejestrowane karty stanowią podstawę wydawnictwa „Informator o zakończonych pracach naukowych i naukowo-badawczych", który od roku 1968 ukazywał się w odstępach kwartalnych. Po zakończeniu kwartału zamykano kompletowanie kart zeszytu „Informatora" i przystępowano do sporządzenia maszynopisu: informacje porządkowano wg UKD, przy czym przy porządkowaniu brano pod uwagę tylko jej pierwszy symbol UKD. Po takim uporządkowaniu kartom dokumentacyjnym nadawano kolejny numer. Następnie na podstawie skróconych tablic UKD wprowadzono nazwy tytułowe działów i poddziałów stosownie do tematyki prac w danym zbiorze kart dokumentacyjnych. Wydawnictwo to uzupełniano opracowywanym ręcznie indeksem przedmiotowym i autorskim oraz spisem treści. Całość opracowanego w ten sposób materiału stanowiła egzemplarz wydawniczy „Informatora".

Ten konwencjonalny, powszechnie do dziś w kraju stosowany sposób opracowywania wydawnictw tego typu powodował spiętrzenia prac po zamknięciu kwartału i wydłużał cykl wydawniczy do 6 i więcej miesięcy, licząc od daty wpływu pierwszych kart nowego kwartału.

---

[1]) W tym celu utworzona została wspólna grupa projektowo-konsultacyjna, w której pracach uczestniczyli ze strony CIINTE: W. Piróg, T. Markowski, a ze strony ZOWARU: A. Targowski, M. Borowska, R. Zając, J. Kalisiak, W. Osińska, przy czym głównym projektantem systemu był R. Zając, w początkowej fazie wdrażania systemu duże zasługi przy konwersji dokumentacji inte położyła A. Rycińska (ZOWAR).

Podstawowe założenia systemu IGA

System IGA w części dotyczącej przygotowywania egzemplarza wydawniczego „Informatora" opiera się na dotychczasowej metodyce pracy z tą różnicą, że:

- pewne części wykonywane są maszynowo
- częstotliwość ukazywania się „Informatora" została skrócona do jednego miesiąca.

Niezależnie od powyższego, użycie elektronicznej maszyny cyfrowej pozwoliło na wprowadzenie szeregu zmian w stosunku do ręcznego, konwencjonalnego opracowywania „Informatora", a mianowicie:

- zmieniono kryterium porządkujące z pierwszego symbolu UKD pracy na pełny symbol UKD. Wprowadziło to konieczność łączenia poszczególnych symboli pracy w jeden symbol długi symbolem relacji „:" (dwukropek). Realizując maszynowe porządkowanie pracy wg UKD zrezygnowano ze skomplikowanego mechanizmu porównań i zastąpiono konwersją symbolu na tzw. notację Freemana [1], [2], pozwalającą na maszynowe porównanie symboli UKD;
- wprowadzono układ odsyłaczy referencyjnych. Odsyłacze te tworzone są z dalszych symboli UKD;
- wprowadzono do „Informatora" abstrakty prac;
- zastosowanie EMC pozwoliło sporządzać roczne indeksy zbiorcze oraz wyszukiwać informację wg kryteriów: symbol UKD, autor, deskryptor, hasło przedmiotowe.

Ogólna organizacja systemu IGA

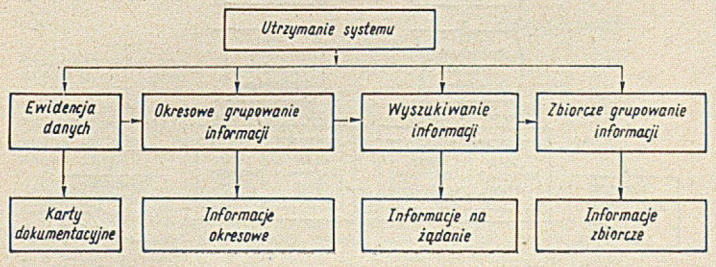
System IGA został zaprogramowany na EMC IBM 1440. Zestaw maszyny obejmuje:

- jednostkę centralną wraz z pamięcią wewnętrzną (12 000 znaków)
- czytnik (perforator kart)
- drukarkę wierszową oraz
- trzy jednostki wymienne dysków magnetycznych.

Z uwagi na znaczną ilość programów (ponad 50), utworzono dyskową bibliotekę programów systemowych i wyposażono system w monitor pozwalający na automatyczne przekazywanie sterowania pomiędzy programami. Ogólnie biorąc, system IGA składa się z 5 jednostek przetwarzania danych (rys. 1), a mianowicie:

- jednostka utrzymania systemu. Główne jej funkcje — to utrzymanie dyskowej biblioteki programów systemowych, włączanie nowych programów, przekazywanie sterowania pomiędzy programami;
- jednostka ewidencji danych. Pozwala na włączenie następujących kartotek dyskowych:
  — kartoteka nazw tytułowych,
  — miesięczna kartoteka kart dokumentacyjnych;
- jednostka okresowego grupowania informacji. Pozwala ona na otrzymanie egzemplarza wydawniczego „Informatora" na dany miesiąc („Informatora" w układzie UKD, łącznie z indeksami, tj. autorskim i przedmiotowym, do zbioru informacji z danego okresu);

Rys. 1. Schemat ideowy systemu IGA



- jednostka wyszukiwania. Pozwala ona na otrzymanie zestawień dokumentacyjnych wg zadanych kryteriów (symbol UKD, autor, hasło przedmiotowe);
- jednostka zbiorczego grupowania informacji. Pozwala ona na kumulowanie zbioru danych i opracowywanie zbiorczego „Informatora" (np. rocznych indeksów).

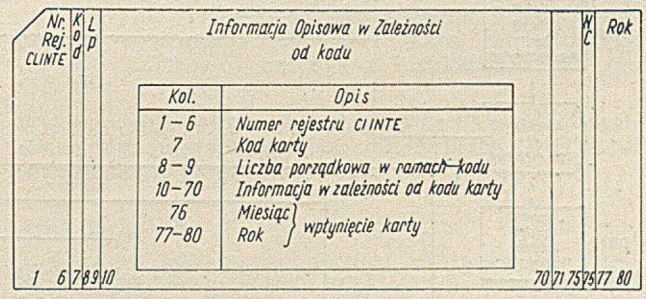
Przygotowanie danych do „Informatora"

Karta dokumentacyjna, po wpłynięciu do CIINTE, podlega kontroli co do poprawności jej budowy i zostają na niej naniesione hasła indeksu przedmiotowego, wyrazy kluczowe oraz numer rejestru. Po wydziurkowaniu na kartach maszynowych dla każdej karty dokumentacyjnej powstaje plik kart dziurkowanych (maszynowych nośników informacji). Budowa pliku jest następująca:

- każda karta pliku zawiera numer rejestru oraz miesiąc i rok wpłynięcia karty;
- wewnątrz pliku karty ułożone są kolejno wg symboli kodowych, jak na tablicy.

Tablica

Symbol kodowy	Rozwój informacji
1	symbol UKD
2	autor
3	tytuł
4	dane bibliograficzne
5	charakter pracy
6	abstrakt
7	deskryptor (hasło przedmiotowe)



Rys. 2. Budowa karty dziurkowanej

W obrębie każdego kodu dziurkowana jest dwucyfrowa liczba porządkowa, co pozwala na wprowadzenie do 99 kart na danym kodzie. Fakt ten oznacza, że karta dokumentacyjna może być scharakteryzowana nie więcej niż 99 symbolami UKD, tyluż autorami i deskryptorami (kody 1, 2, 7). Poza tym karta dziurkowana nakłada ograniczenie na długość pojedynczego symbolu UKD, autora, deskryptora — maks. 61 znaków. W przypadku kodów: 3, 4, 5, 6 oznacza to, że informacja opisowa, np. abstrakt, może być kontynuowana na kolejnych kartach (maks. 99 kart) po 61 znaków na karcie.

Budowa karty dziurkowanej przedstawiona jest na rys. 2.

Karty dziurkowane ułożone wg numeru rejestracyjnego CIINTE kodu i liczby porządkowej w ramach kodu po sprawdzeniu i naniesieniu poprawek tworzą podstawowy zbiór informacji dla systemu.

Rys. 3 przedstawia kartę dokumentacyjną pracy.

66.067.1 Sączenie. Rafinacja 64,129:IMN, MPC
622.335 Górnictwo miedzi 11

MARCZYŃSKI W.: Sposób filtracji koncentratów rud miedzi. Monografia. Instytut Metali Nieżelaznych. Gliwice. 1964. Ss.42. Rys.16. Tabl.4. Wykr.6.

Zgłoszono do Urzędu Patentowego PRL dnia 5 maja 1964 r.

Przeprowadzono badania w skali laboratoryjnej i 1/4 technicznej na filtracją koncentratów miedziowych w bębnowym filtrze próżniowym. Określono opór właściwy osadu powstającego w wyniku filtracji oraz jego wpływ na wyniki pracy filtra. Wskazano przyczyny trudnych warunków filtracji koncentratu miedziowego. Wykonano próby zastosowania odczynników powierzchniowych czynnych i flokulatorów w procesie filtracji oraz opisano sposób filtracji.

miedź - koncentraty
filtracja koncentratów miedziowych
rudy miedzi - oczyszczanie

d.c.

Rys. 3. Karta dokumentacyjna

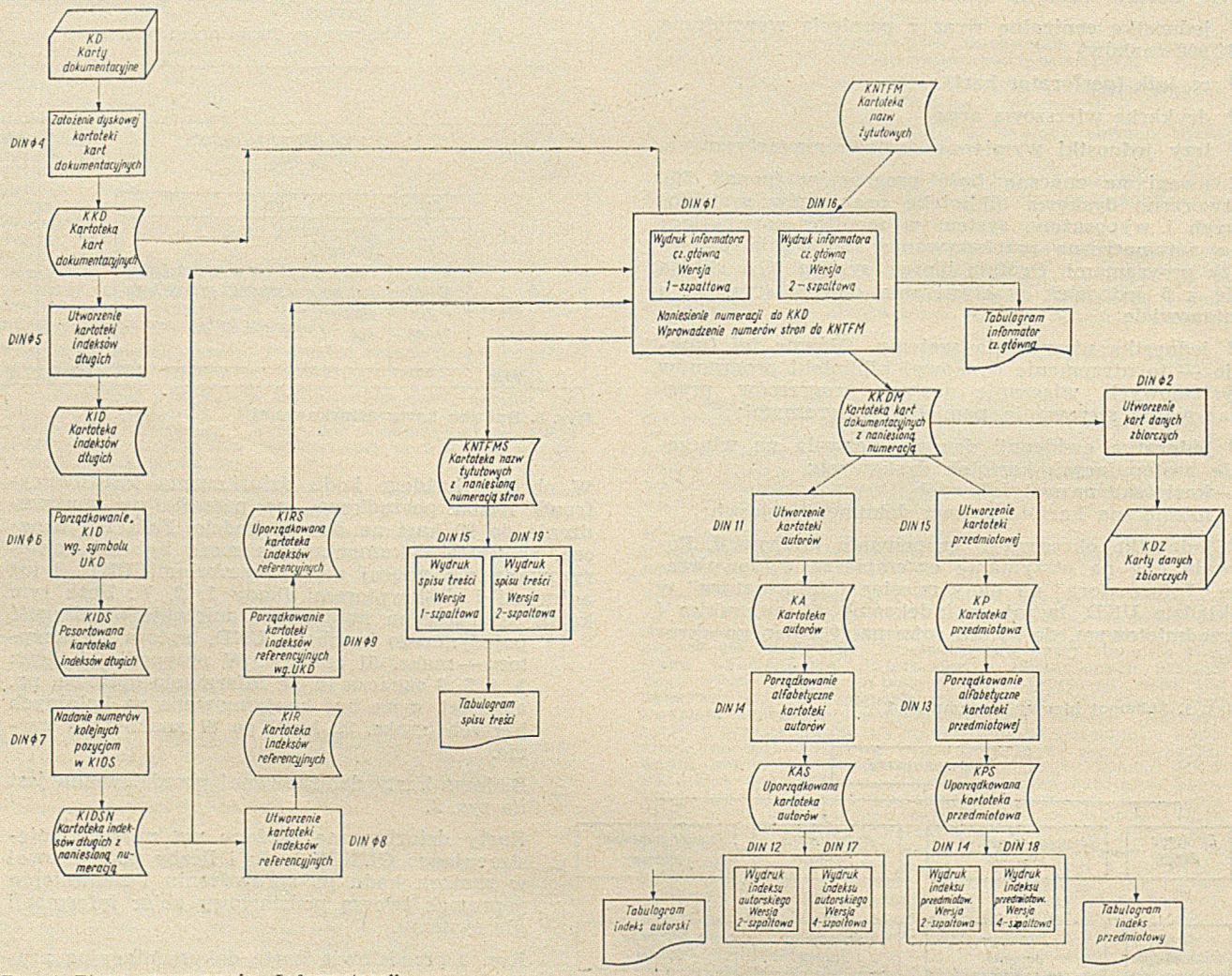
funkcje i czynności wykonywane kolejno w trakcie przetwarzania. Należy tu tylko zaznaczyć, że tego typu organizacja przetwarzania była możliwa dzięki posiadaniu pamięci zewnętrznej o dostępie przypadkowym (dyski magnetyczne).

Postać informacji na wyjściu z maszyny w systemie IGA

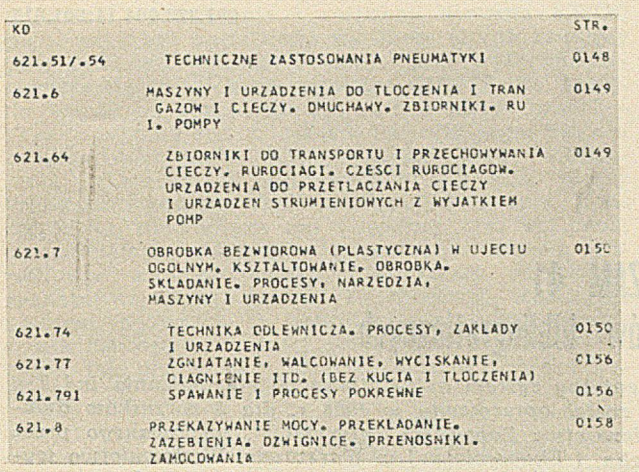
Jak już zostało powyżej zaznaczone, w chwili obecnej nie wszystkie jednostki przetwarzania są w pełni oprogramowane, względnie te nie wszystkie jednostki przetwarzania mogą być realizowane, ponieważ istniejące środki techniczne są ograniczone (brak jest dostatecznej ilości dysków w dyspozycji ZOWAR - Warszawa) w tych samych warunkach jedynie jednostka związana z okresowym grupowaniem informacji jest eksploatowana, czego rezultatem jest „Informator o zakończonych pracach naukowych i naukowo-badawczych". Poszczególne, charakterystyczne strony „Informatora" są przedstawione na rysunkach od 5 do 8.

Maszynowe opracowanie informatora

Na rys. 4 przedstawiono kolejno etapy opracowania „Informatora" o zakończonych pracach naukowych i naukowo badawczych" CIINTE. W niniejszym artykule pominięto opis charakteryzujący poszczególne

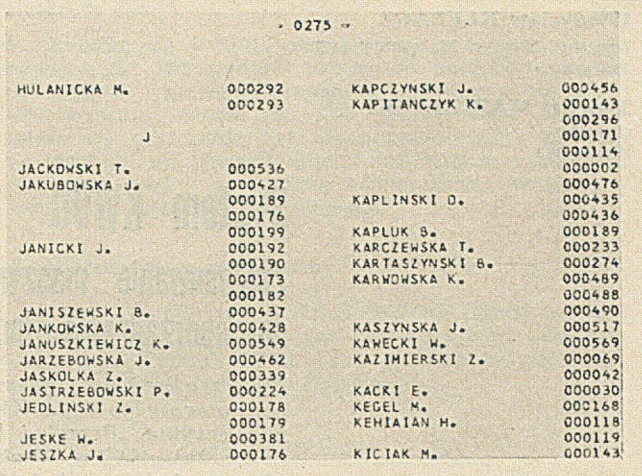


Rys. 4. Etapy opracowania „Informatora"



KD		STR.
621.51/.54	TECHNICZNE ZASTOSOWANIA PNEUMATYKI	0148
621.6	MASZYNY I URZADZENIA DO TLOCZENIA I TRAN GAZOW I CIECZY. DMUCHAWY. ZBIORNIKI. RU I. POMPY	0149
621.64	ZBIORNIKI DO TRANSPORTU I PRZECHOWYWANIA CIECZY. RUROCIAGI. CZESCI RUROCIAGOW. URZADZENIA DO PRZETLACZANIA CIECZY I URZADZEN STRUMIENIOWYCH Z WYJATKIEM POMP	0149
621.7	OBROBKA BEZWIOROWA (PLASTYCZNA) W UJECIU OGOLNYM. KSZTALTOWANIE. OBROBKA. SKLADANIE. PROCESY. NARZEDZIA. MASZYNY I URZADZENIA	0150
621.74	TECHNIKA ODLEWNICZA. PROCESY, ZAKLADY I URZADZENIA	0150
621.77	ZGNIATANIE, WALCOWANIE, WYCISKANIE, CIAGNIENIE ITD. (BEZ KUCIA I TLOCZENIA)	0156
621.791	SPAWANIE I PROCESY POKREWNE	0156
621.8	PRZEKAZYWANIE MOCY. PRZEKLADANIE. ZAZEBIENIA. DZWIGNICE. PRZENOSNIKI. ZAMOCOWANIA	0158

Rys. 5. Fragment strony spisu treści



- 0275 -

HULANICKA M.	000292	KAPCZYNSKI J.	000456
	000293	KAPITANCZYK K.	000143
			000296
J			000171
			000114
JACKOWSKI T.	000536		000002
JAKUBOWSKA J.	000427		000476
	000189	KAPLINSKI O.	000435
	000176		000436
	000199	KAPLUK B.	000189
JANICKI J.	000192	KARCZEWSKA T.	000233
	000190	KARTASZYNSKI B.	000274
	000173	KARWOWSKA K.	000489
	000182		000488
JANISZEWSKI B.	000437		000490
JANKOWSKA K.	000428	KASZYNSKA J.	000517
JANUSZKIEWICZ K.	000549	KAWECKI W.	000569
JARZEBOWSKA J.	000462	KAZIMIERSKI Z.	000069
JASKOLKA Z.	000339		000042
JASTRZEBOWSKI P.	000224	KACKI E.	000030
JEDLINSKI Z.	000178	KEGEL M.	000168
	000179	KEHIAIAN H.	000118
JESKE W.	000381		000119
JESZKA J.	000176	KICIAK M.	000143

Rys. 7. Fragment strony indeksu autorskiego

Efekty operacyjno-ekonomiczne systemu

Wdrożenie opisanego systemu przyniosło szereg efektów operacyjnych i ekonomicznych, z których do najważniejszych należą:

- przerzucenie prac związanych z opracowaniem „Informatora" na EMC i odciążenie personelu technicznego od prac mechanicznych;
- skrócenie czasu opracowania „Informatora" do jednego dnia i to w terminach ściśle określonych;
- możliwość rozszerzenia „Informatora" o dodatkowe kryteria informacyjne (w perspektywie);
- możliwość objęcia tym systemem opracowania dalszych informatorów o podobnej tematyce. W takim przypadku efekty ekonomiczne systemu wystąpiłyby z większą wyrazistością;
- możliwość skompletowania „Informatora" (np. tworzenie scentralizowanego „Informatora", tworzenie „Informatora" rocznego, indeksów rocznych itp.).

Uwagi końcowe

W ramach tej pracy natrafiono na duże trudności z powodu braku polskiego alfabetu. Maszyna IBM 1440, na której opracowywany jest „Informator" posiada drukarkę wierszową z dużymi literami alfabetu łacińskiego, cyframi i znakami specjalnymi. Wydruk tekstu wymaga więc korekt typu dostawiania znaków diakrytycznych do pewnych liter, naturalnie w miarę konieczności. Chodzi tu o znaczki przy literach ę, ć, ą, ł, ń, ó, ś, ź. Należy jednak podkreślić, że w indeksie autorskim i rzeczowym nie występują żadne naruszenia alfabetycznego porządku polskiego alfabetu. Korekty natomiast są wymagane przy użyciu symboli matematycznych i liter alfabetów niełacińskich.

Wydawanie „Informatora"

Stronice drukowane przez drukarkę współpracującą z maszyną IBM 1440 są przekazywane do Zakładu Usług Dokumentacyjnych CIINTE i za pośrednictwem matrycy wykonanej na kserografie są powielane w dowolnej ilości i wydawane w formie zwartej.

Do przetworzenia w 1968 r. weszło ok. 5000 informacji w postaci kart dokumentacyjnych. Ten zbiór informacji został zgrupowany automatycznie w 11 numerach „Informatora". Zostało wydziurkowanych w związku z tym ok. 60 000 kart maszynowych 80-kolumnowych zawierających ok. 3 000 000 znaków alfa-numerycznych. Sporządzenie jednego numeru „Informatora", tj. odpowiednie zgrupowanie informacji na maszynie łącznie ze sporządzeniem indeksów i pełnym wydrukiem jednego numeru zawierającego ok. 600 informacji (łącznie z indeksami) trwa ok. 4 godzin.

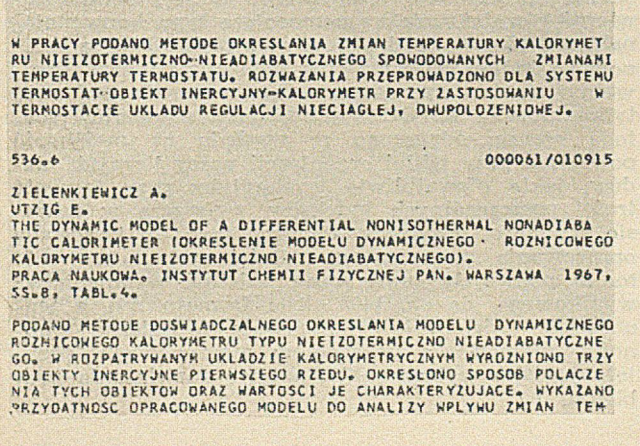
Wykonywanie tej pracy stanowi jeden z zasadniczych elementów nabywania doświadczeń przez CIINTE w tej dziedzinie przetwarzania.

BIBLIOGRAFIA

[1] Robert R. Freeman and Paolinte Atherton — File Organisation and Search Strategy using UDC in Mechanized Reference Retrieval Systems — Preprint of a Paper for Presentation at the FID/IFIP Conference on Mechanized Information Storage Retrieval and Dissemination, Rome. June, 15, 1967.

[2] Robert R. Freeman — Computers and Classification Systems — Journal of Documentation, 20/3, 137—145 (1964).

Rys. 6. Fragment strony „Informatora" właściwego



- 0031 -

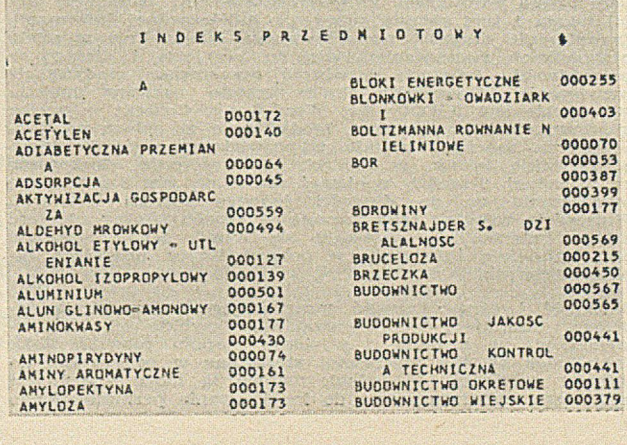
W PRACY PODANO METODE OKRESLANIA ZMIAN TEMPERATURY KALORYMET RU NIEIZOTERMICZNO-NIEADIABATYCZNEGO SPOWODOWANYCH ZMIANAMI TEMPERATURY TERMOSTATU. ROZWAZANIA PRZEPROWADZONO DLA SYSTEMU TERMOSTAT-OBIEKT INERCYJNY-KALORYMETR PRZY ZASTOSOWANIU W TERMOSTACIE UKLADU REGULACJI NIECIAGLEJ, DWUPOLOZENIOWEJ.

536.6 000061/010915

ZIELENKIEWICZ A.
UTZIG E.
THE DYNAMIC MODEL OF A DIFFERENTIAL NONISOTHERMAL NONADIABA TIC CALORIMETER (OKRESLENIE MODELU DYNAMICZNEGO ROZNICOWEGO KALORYMETRU NIEIZOTERMICZNO NIEADIABATYCZNEGO).
PRACA NAUKOWA. INSTYTUT CHEMII FIZYCZNEJ PAN. WARSZAWA 1967, SS.8, TABL.4.

PODANO METODE DOSWIADCZALNEGO OKRESLANIA MODELU DYNAMICZNEGO ROZNICOWEGO KALORYMETRU TYPU NIEIZOTERMICZNO NIEADIABATYCZNE GO. W ROZPATRYWANYM UKLADZIE KALORYMETRYCZNYM WYROZNIONO TRZY OBIEKTY INERCYJNE PIERWSZEGO RZEDU. OKRESLONO SPOSOB POLACZE NIA TYCH OBIEKTOW ORAZ WARTOSCI JE CHARAKTERYZUJACE. WYKAZANO PRZYDATNOSC OPRACOWANEGO MODELU DO ANALIZY WPLYWU ZMIAN TEM

Rys. 8. Fragment strony indeksu przedmiotowego



- 0283 -

INDEKS PRZEDMIOTOWY

A		BLOKI ENERGETYCZNE	000255
		BLONKOWKI - OWADZIARK I	000403
ACETAL	000172		
ACETYLEN	000140	BOLTZMANNA ROWNANIE N IELINIOWE	000070
ADIABETYCZNA PRZEMIAN A	000064	BOR	000053
ADSORPCJA	000045		000387
AKTYWIZACJA GOSPODARC ZA	000559	BOROWINY	000399
ALDEHYD MROWKOWY	000494	BRETSZNAJDER S. DZI ALALNOSC	000177
ALKOHOL ETYLOWY - UTL ENIANIE	000127	BRUCELOZA	000569
ALKOHOL IZOPROPYLOWY	000139	BRZECZKA	000215
ALUMINIUM	000501	BUDOWNICTWO	000450
ALUN GLINOWO-AMONOWY	000167		000567
AMINOKWASY	000177	BUDOWNICTWO JAKOSC PRODUKCJI	000565
	000430	BUDOWNICTWO KONTROL A TECHNICZNA	000441
AMINOPIRYDYNY	000074	BUDOWNICTWO OKRETOWE	000441
AMINY AROMATYCZNE	000161	BUDOWNICTWO WIEJSKIE	000111
AMYLOPEKTYNA	000173		000379
AMYLOZA	000173		

Władysław KLEPACZ
Instytut Maszyn Matematycznych
Warszawa

Tadeusz MARKOWSKI
CIINTE
Warszawa

681.322.004.14:001.815

System KWOC

Zastosowanie maszyny ZAM 41 do sporządzania indeksów bibliograficznych

Pierwszy w Polsce eksperymentalny system automatycznego sporządzania indeksu permutacyjnego typu KWOC został opracowany w 1968 r. dla 5 roczników dwumiesięcznika „Przegląd Piśmiennictwa Zagadnień Informacji", wydawanego przez Centr. Instytut Inf. Nauk. Techn. i Ekonomicznej w Warszawie. Wydawnictwo wynikowe obejmuje około 5000 pozycji bibliograficznych w postaci: 1) podstawowego wykazu bibliograficznego uporządkowanego wg UKD, 2) indeksu permutacyjnego typu KWOC tytułów publikacji (w języku polskim), 3) indeksu autorskiego, 4) indeksu czasopism. Wszystkie postaci wydawnictw są powiązane ze sobą za pośrednictwem odsyłaczy. W artykule opisano organizację systemu, sposób wprowadzania danych i programy na maszynę ZAM-41. W systemie zastosowano metodę automatycznego sprowadzania wyrazów kluczowych do jednolitej formy gramatycznej.

1. Wprowadzenie

Wśród licznych użytkowych prób wykorzystania elektronicznych maszyn cyfrowych do zautomatyzowania procesów informacji naukowo-technicznej najbardziej rozpowszechnionym zastosowaniem jest w chwili obecnej sporządzanie indeksów permutacyjnych do przeglądów bibliograficznych.

Indeksy permutacyjne znane były już od dawna w technice katalogowania zbiorów bibliograficznych jako bardzo efektywne uzupełnienie konwencjonalnych indeksów przedmiotowych i autorskich, jednak z uwagi na szczególnie wysoką ich pracochłonność realizowane były w praktyce bardzo rzadko. Stąd też spopularyzowanie metody indeksowania permutacyjnego nastąpiło dopiero z chwilą zastosowania do tego celu elektronicznych maszyn cyfrowych. Aktualnie przeważająca część znanych światowych przeglądów bibliograficznych wyposażona jest w indeksy permutacyjne, które są tam wyeksponowane jako podstawowe narzędzie wyszukiwania informacji.

Pierwsza praktyczna próba w tym zakresie została zrealizowana w USA już w roku 1958 na maszynie IBM 709. System ten został nazwany przez jego twórcę H. P. LUHNA „KWIC" (Key-Word-In-Context = wyraz kluczowy w kontekście). Nazwa ta przyjęła się powszechnie jako pojęcie o charakterze ogólnym („indeksy typu KWIC") dla określania zautomatyzowanego sporządzania indeksów permutacyjnych, pomimo opracowania w międzyczasie licznych modyfikacji rozwiązania pierwotnego. W rozwiązaniu tym, określanym jako klasyczne, wyrazy znaczące wyeksponowane są w stałym miejscu środkowej części wiersza drukarskiego. Powoduje to konieczność rozbijania każdego tekstu (np. tytułu publikacji) na dwie części. Część zlokalizowana po prawej stronie wyrazu kluczowego jest kontynuacją logiczną treści tekstu, następującej bezpośrednio po tym wyrazie, natomiast część tekstu zlokalizowana po jego lewej stronie stanowi zakończenie treści całego tekstu. Celem wyróżnienia wyrazów kluczowych w tekście są one wyrównywane i oddzielane od lewej części tekstu wyraźnym światłem w każdym wierszu. W rozwiązaniu klasycznym metody KWIC przeznacza się na jeden wyraz kluczowy wraz z kontekstem tylko jeden wiersz drukarki maszyny cyfrowej, co powoduje automatyczne skracanie przez maszynę tekstów przekraczających maksymalna długość wiersza (ucinanie lewej, tzn. końcowej części tekstu). Operacja taka często powodować może zniekształcenie istniejącej treści tekstu i tym samym zubożenie indeksowanej informacji. Ograniczenie długości permutowanego tekstu było w wielu przypadkach rozwiązaniem nie do przyjęcia i dlatego w poszukiwaniu lepszych rozwiązań opracowano modyfikację rozwiązania klasycznego, które zostało nazwane metodą albo systemem KWOC (Key-Word-Out-of-Context = wyraz kluczowy poza kontekstem). Modyfikacja ta polega na tym, że wyraz znaczący zostaje wyprowadzony poza kontekst, a wszystkie teksty, w których on występuje, drukowane są w pełnej długości, ponieważ zniesione jest wspomniane ograniczenie co do długości jednego wiersza drukarki. Wprawdzie wskutek tego wzrasta w sposób istotny objętość całego indeksu, ale uzyskuje się lepszą jego czytelność oraz pełniejszą informację o poszczególnych pozycjach bibliograficznych.

Szybkość i ekonomiczność maszynowego sporządzania indeksów permutacyjnych spowodowały szybkie ich rozpowszechnienie w krajach o wysokim stopniu rozwoju ETO. Jakkolwiek są one najczęściej stosowane w odniesieniu do tytułów publikacji, to jednak metoda ta może być i jest już często wykorzystywana do indeksowania permutacyjnego również innych rodzajów tekstów informacyjnych (streszczenia, opisy deskryptorowe itp.).

Centralny Instytut Informacji Naukowo-Technicznej i Ekonomicznej (CIINTE) podjął w IV. kwartale 1967 roku prace eksperymentalne, zmierzające do opracowania indeksu permutacyjnego przy użyciu elektronicznej maszyny cyfrowej.

W wyniku przeprowadzonych prac studialnych podjęto decyzję realizacji systemu na polskiej maszynie ZAM 41 oraz rozwiązanie indeksu metodą KWOC. Decyzja wyboru maszyny ZAM 41 była uzasadniona tym, że będzie ona na przestrzeni najbliższych lat najbardziej rozpowszechnionym w kraju modelem maszyny do przetwarzania danych (co stwarza warunki szerszego rozpowszechnienia systemu u innych użytkowników), jak również możliwością skorzystania z doświadczeń i uzyskania merytorycznej pomocy ze strony Instytutu Maszyn Matematycznych, który opracował i wdrożył na tej maszynie pierwszy w kraju system automatyzacji informacji bibliograficznych (system INBI).

2. Przedmiot oraz zakres systemu

Przedmiotem systemu było 5 roczników dwumiesięcznika „Przegląd Piśmiennictwa Zagadnień Informacji" wydawanego przez CIINTE. Obejmowały one okres lat 1962—1966 o łącznej objętości powyżej 5000 pozycji bibliograficznych. Jakkolwiek podstawowym celem systemu nazwanego roboczo Systemem KWOC, miało być automatyczne sporządzenie indeksu typu KWOC, to jednak uznano, że indeks taki nie może być odrębnym wydawnictwem, lecz powinien zostać opracowany w bezpośrednim powiązaniu ze zbiorczym wykazem bibliograficznym za cały wspomniany okres. Oprócz tego wydawnictwo takie powinno zawierać również indeksy pomocnicze typu konwencjonalnego, stosowane dotąd w zeszytach PPZI, a mianowicie indeks autorski oraz indeks czasopism. Pierwotny zamiar opracowania na maszynie cyfrowej tylko indeksu KWOC a pozostałych elementów metodą ręczną odrzucono ze względu na możliwość powstania przy takiej organizacji pracy licznych błędów, jakie niewątpliwie wystąpiłyby przy łączeniu treści poszczególnych indeksów, zwłaszcza w zakresie zgodności odsyłaczy. Pomimo tego, że koncepcja opracowania maszynowego całości systemu co najmniej podwajała sumaryczną objętość wydawnictwa wynikowego (do ok. 1400 stron formatu A4), a jednocześnie wielokrotnie zwiększała techniczne trudności opracowania programów maszynowych, przyjęto ją jako wyłączną alternatywę rozwiązania problemu.

3. Założenia organizacyjne realizacji systemu

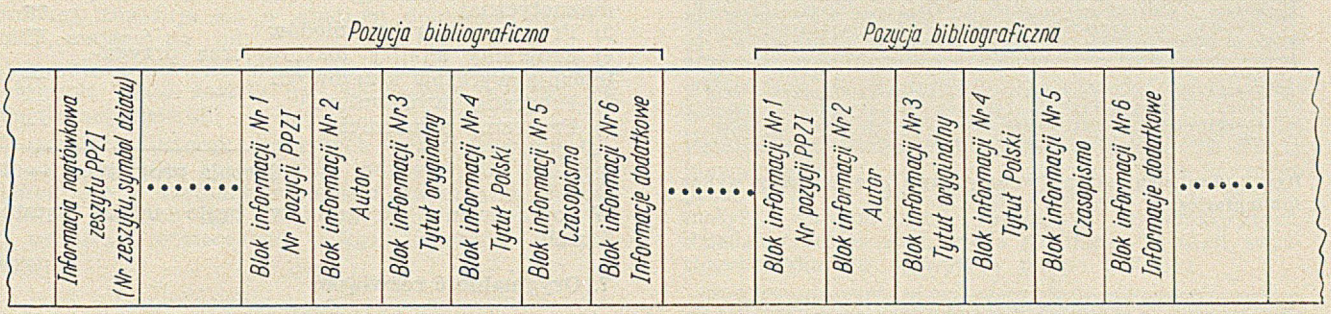
Czasokres opracowania i realizacji systemu KWOC planowany był początkowo na okres ok. 10 miesięcy przy założeniu wykonania całości prac wyłącznie systemem zlecenia ich na zewnątrz CIINTE. Ze względu jednak na bezpośrednie oraz perspektywiczne korzyści, wynikające z posiadania własnego zespołu programistów, postanowiono przyjąć zasadę wyszkolenia odpowiednich pracowników CIINTE, którzy wykonaliby znaczną część prac programowania. Koncepcja ta ze względu na konieczny czas szkolenia oraz mniejszą wydajność wspomnianego zespołu, spowodować musiała jednak rozszerzenie czasokresu realizacji systemu do ok. 18 miesięcy. Realizacja systemu wykazała, że wskutek poważnego rozszerzenia zakresu oraz ilości programów maszynowych, nawet tak przedłużony termin okazał się bardzo napięty.

Sposób zapisu pozycji bibliograficznych oraz poszczególnych bloków informacji na taśmie dziurkowanej ilustruje rys. 1.

Zakładając, że średnio jedna pozycja bibliograficzna zawiera ok. 250 znaków pisarskich (maksymalnie do 500 znaków), system wymagał przeniesienia na taśmę dziurkowaną łącznie ok. 1,25 mln znaków. Stopień trudności perforacji był stosunkowo duży, głównie ze względu na znaczny udział informacji w różnych językach obcych oraz dużą różnorodność wariantów struktury opisu bibliograficznego.

5. Wydawnictwa wynikowe

Jak już wspomniano, System KWOC składa się z 4 następujących wydawnictw wynikowych, powiązanych z zeszytami PPZI oraz wzajemnie pomiędzy sobą za pośrednictwem odsyłaczy:



Rys. 1

Jako maszynowy nośnik informacji przyjęto 5-ścieżkową taśmę dziurkowaną, która przy dokumentach o objętości powyżej 80 czy 90 znaków wykazuje wyraźną przewagę nad kartami dziurkowanymi. Jako metodę kontroli poprawności przenoszenia informacji na taśmę dziurkowaną przyjęto sczytywanie tabulogramów. Założono wreszcie, że wydawnictwa wynikowe z maszyny cyfrowej zostaną powielone w ilości kilkuset egzemplarzy metodą kserograficzną w formacie A4 (rozmiar zeszytów PPZI).

Należy podkreślić, że tego rodzaju rozmiary jednorazowego automatycznego indeksowania są nadal rzadkością nawet w skali światowej. Przytłaczająca większość maszynowych opracowań indeksów permutacyjnych ogranicza się maksymalnie do liczby ok. 1000 pozycji bibliograficznych. Pod tym względem omawiany system wysuwa Polskę zdecydowanie na pierwsze miejsce wśród krajów socjalistycznych (radziecki indeks typu KWIC do czasopisma „Refieratiwnyj Żurnał — Naucznaja i Tiechniczeskaja Informacja" obejmuje jednorazowo do ok. 300 pozycji bibliograficznych). Warto również podkreślić, że rozwiązanie typu KWOC, który stanowi bardziej rozwiniętą a jednocześnie mniej opisaną w literaturze metodę indeksowania permutacyjnego, nie było jeszcze zrealizowane w krajach naszego obozu.

4. Dane wejściowe

Podstawowym dokumentem wejściowym systemu jest pozycja bibliograficzna z zeszytów PPZI. Pozycja ta przenoszona jest na taśmę dziurkowaną z zachowaniem jej wewnętrznej struktury, eliminując jedynie informację o formacie wydawnictwa oraz analizę (streszczenie) publikacji. W strukturze tej wyodrębniono poprzez zastosowanie separatorów (znaków oddzielających) następujące bloki informacji:

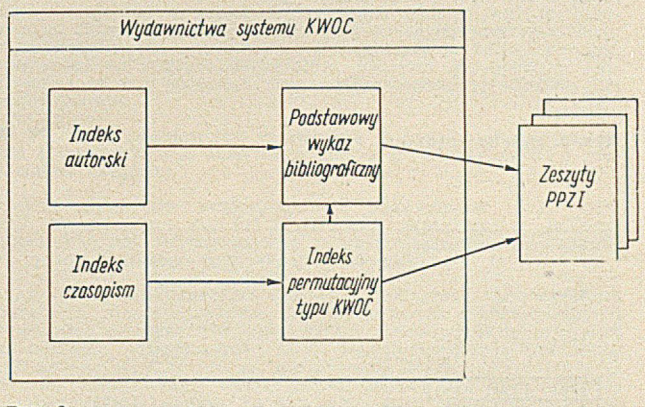
1) numer bieżący pozycji w zeszycie PPZI,
2) nazwisko autora (autorów) publikacji,
3) tytuł oryginalny publikacji,
4) polskie tłumaczenie obcojęzycznego tytułu publikacji,
5) skrót czasopisma,
6) dodatkowe informacje bibliograficzne (miejsce i rok wydania, nr czasopisma, objętość tekstu publikacji itp.).

1) podstawowego wykazu bibliograficznego, obejmującego skumulowane pozycje bibliograficzne wg układu stosowanego dotąd w zeszytach PPZI (uporządkowanie wg kryterium Uniwersalnej Klasyfikacji Dziesiętnej). Pozycje te powiązane są z zeszytami PPZI za pośrednictwem szczegółowych odsyłaczy (nr pozycji, nr zeszytu, rok);

2) indeksu permutacyjnego typu KWOC polskich tytułów publikacji (oryginalnych i tłumaczonych tytułów obcojęzycznych), powiązanych odsyłaczami z wyżej wymienionym podstawowym wykazem bibliograficznym;

3) indeksu autorskiego, powiązanego odsyłaczami z podstawowym wykazem bibliograficznym;

4) indeksu czasopism, powiązanego odsyłaczami z podstawowym wykazem bibliograficznym.

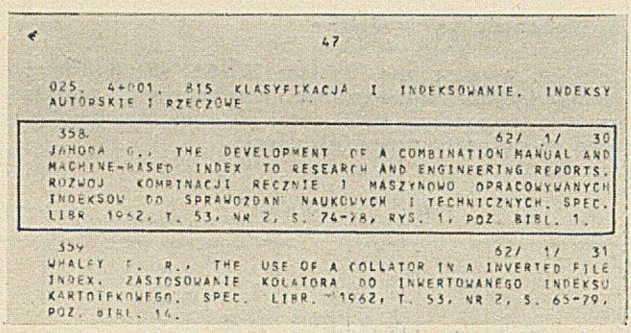


Rys. 2

Rys. 2 ilustruje wzajemne powiązanie poszczególnych wydawnictw za pośrednictwem odsyłaczy, natomiast rys. 3—6 układ typograficzny poszczególnych wydawnictw wynikowych.

Rzeczywista objętość tekstu wydawnictw, uzyskana w próbnej fazie realizacji systemu wyniosła:

1) podstawowy wykaz bibliograficzny	— 517	stron formatu A4,
2) indeks KWOC	— 720	„ „ „
3) indeks autorski	— 147	„ „ „
4) indeks czasopism	— 26	„ „ „
łącznie	1410	„ „ „



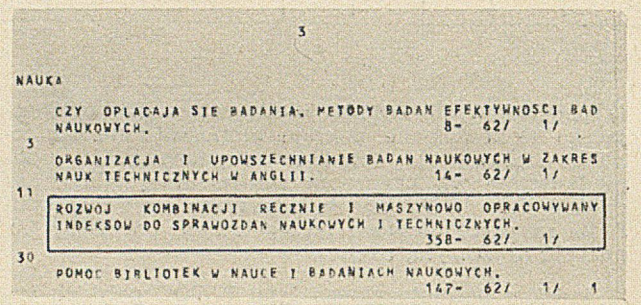
47

025.4+001.815 KLASYFIKACJA I INDEKSOWANIE. INDEKSY AUTORSKIE I RZECZOWE

358 62/ 1/ 30
JAHODA G., THE DEVELOPMENT OF A COMBINATION MANUAL AND MACHINE-BASED INDEX TO RESEARCH AND ENGINEERING REPORTS. ROZWOJ KOMBINACJI RECZNIE I MASZYNOWO OPRACOWYWANYCH INDEKSOW DO SPRAWOZDAN NAUKOWYCH I TECHNICZNYCH. SPEC. LIBR. 1962, T. 53, NR 2, S. 74-78, RYS. 1, POZ. BIBL. 1.

359 62/ 1/ 31
WHALEY F. R., THE USE OF A COLLATOR IN A INVERTED FILE INDEX. ZASTOSOWANIE KOLATORA DO ODWROTNEGO INDEKSU KARTOTEKOWEGO. SPEC. LIBR. 1962, T. 53, NR 2, S. 65-79, POZ. BIBL. 14.

Rys. 3. Fragment strony wydawnictwa „Podstawowy wykaz bibliograficzny"

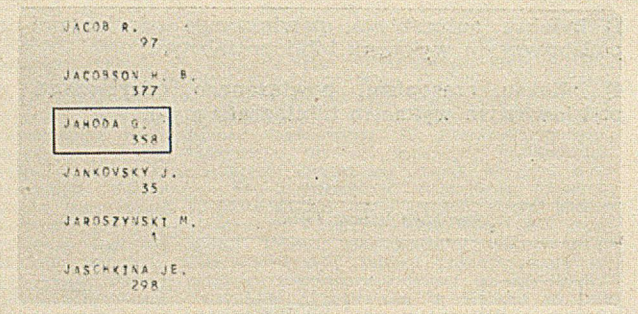


3

NAUKA

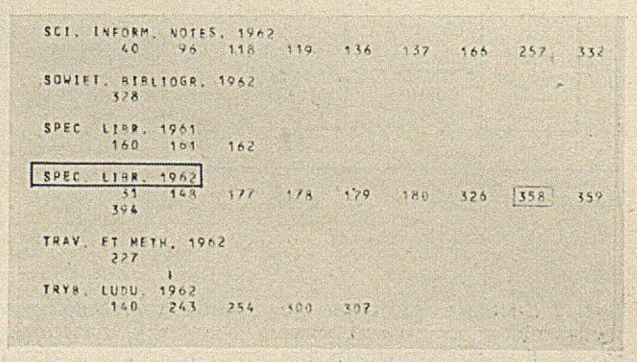
CZY OPLACAJA SIE BADANIA, METODY BADAN EFEKTYWNOSCI BAD NAUKOWYCH. 8- 62/ 1/
3
ORGANIZACJA I UPOWSZECHNIANIE BADAN NAUKOWYCH W ZAKRES NAUK TECHNICZNYCH W ANGLII. 14- 62/ 1/
11
ROZWOJ KOMBINACJI RECZNIE I MASZYNOWO OPRACOWYWANY INDEKSOW DO SPRAWOZDAN NAUKOWYCH I TECHNICZNYCH. 358- 62/ 1/
30
POMOC BIBLIOTEK W NAUCE I BADANIACH NAUKOWYCH. 147- 62/ 1/ 1

Rys. 4. Indeks permutacyjny KWOC (fragment)



JACOB R.
97
JACOBSON H. B.
377
JAHODA G.
358
JANKOVSKY J.
35
JAROSZYNSKI M.
1
JASCHKINA JE.
298

Rys. 5. Indeks autorski



SCI. INFORM. NOTES. 1962
40 96 118 119 136 137 166 257 332
SOWIET. BIBLIOGR. 1962
328
SPEC. LIBR. 1961
160 161 162
SPEC. LIBR. 1962
51 148 177 178 179 180 326 358 359
394
TRAV. ET METH. 1962
227
TRYB. LUDU 1962
140 243 254 300 307

Rys. 6. Indeks czasopism (fragment)

6. Programy

Wg pierwotnych szacunków system miał obejmować ok. 10 programów o łącznej pojemności ok. 10 000 rozkazów maszyny. W wyniku licznych modyfikacji w trakcie realizacji systemu, dających w rezultacie rozszerzenie zakresu czynności przejmowanych przez maszynę, system rozszerzył się do 18 programów o pojemności powyżej 20 000 rozkazów maszyny. Jako język programowania użyty został Podstawowy Język Programowania (PJES), który w momencie przystąpienia do programowania był jedynym przydatnym do tego celu elementem oprogramowania maszyny ZAM 41.

Aktualnie system składa się z następujących programów:

1) Programy kontroli danych	— 1
2) Programy zakładania i modyfikacji ewidencji pozycji bibliograficznych na taśmie magnetycznej	— 5
3) Programy korekty błędów	— 2
4) Programy analizy tekstów oraz przygotowania wyrazów kluczowych	— 4
5) Programy wydawnicze	— 5
6) Programy pomocnicze	— 1
łącznie programów	— 18

W opracowaniu programów brało udział łącznie 6 osób.

7. Oryginalność rozwiązań

Opisany system zawiera szereg rozwiązań oryginalnych, rozszerzających konwencjonalną koncepcję indeksu typu KWOC. W tym zakresie należy wymienić w pierwszym rzędzie następujące istotne ulepszenia w stosunku do spotykanych w literaturze wzorców:

1) sprowadzenie wyrazów kluczowych do jednolitej formy gramatycznej,

2) możliwość uwzględnienia synonimów wyrazów kluczowych.

Powyższe cele osiągnięto metodą automatycznego kojarzenia przez maszynę wytypowanych wyrazów kluczowych z występującymi w tytułach mutacjami gramatycznymi tych wyrazów. Metoda ta stanowi znaczne ułatwienie oraz przyspieszenie prac związanych z przyporządkowaniem poszczególnym wyrazom kluczowym odpowiadających im wyrazów tekstowych, które to czynności są szczególnie pracochłonne i żmudne w wykonaniu ręcznym.

8. Wyposażenie maszynowe

System zrealizowano na standardowym zestawie maszyny ZAM 41. Zestaw minimalny maszyny, zapewniający działanie wszystkich programów, powinien posiadać następującą charakterystykę:

1) jeden czytnik taśmy dziurkowanej
2) jedna dziurkarka taśmy papierowej
3) 12K słów pamięci operacyjnej
4) 64K słów pamięci bębnowej
5) cztery jednostki pamięci taśmowej
6) jedna drukarka wierszowa
7) monitor dalekopisowy.

9. Uwagi końcowe

Opracowanie opisanego systemu, oprócz efektów bezpośrednich, pozwoliło uzyskać szereg dodatkowych korzyści, a mianowicie:

1) statystyczne udokumentowanie częstotliwości występowania wszystkich wyrazów w tytułach publikacji, co stanowić będzie istotną pomoc przy opracowywaniu tezaurusa dla dziedziny zagadnień informacji,

2) dokładniejsze poznanie specyfiki języka polskiego pod względem jego właściwości dla celów automatycznego przetwarzania tekstów,

3) wyszkolenie zespołu pracowników wyspecjalizowanych w problematyce praktycznego zastosowania

elektronicznych maszyn cyfrowych do automatyzacji procesów informacyjnych,

4) zapoznanie pracowników Działowego Ośrodka Zagadnień Informacji z istotnymi aspektami automatyzacji indeksowania,

5) wypróbowanie oprogramowania oraz działania poszczególnych modułów prototypu maszyny ZAM 41 w użytkowym systemie złożonego przetwarzania dużych zbiorów informacji tekstowych.

W opracowaniu Systemu oprócz autorów niniejszej publikacji brali udział następujący pracownicy:
— ze strony Instytutu Maszyn Matematycznych:
H. Kozłowska, A. Makowski, J. Wierzbowski, M. Wójcicka;
— ze strony CIINTE:
J. Choroszuszyna, T. Cieszewska, J. Pawłowicz, J. Rowińska, E. Ścibor, A. Targowska, F. Żeromska.
W chwili obecnej System KWOC, po opracowaniu na maszynie ZAM 41 w listopadzie 1968 r. wersji próbnej, znajduje się w stadium dokładnej weryfikacji, mającej na celu:

1) usunięcie ewentualnych błędów w danych wejściowych, jakie zostały nie zauważone podczas ich pierwotnej kontroli;

2) usunięcie usterek wykrytych przez maszynę cyfrową w wyniku opracowania indeksów, jak np. niejednolitość w skrótach czasopism, nadmierna długość tytułów polskich itp.;
3) ustalenie ostatecznego wykazu wyrazów kluczowych.

Po wykonaniu wyżej wymienionych prac sporządzona zostanie w kwietniu 1969 r. ostateczna wersja indeksów, przeznaczona do opublikowania i rozpowszechnienia.

Zakłada się, że opisany system będzie w przyszłości wykorzystany przez CIINTE do indeksowania następnych roczników PPZI. Założone szerokie granice pojemności poszczególnych bloków informacji w danych wejściowych pozwalają również na bezpośrednią adaptację całego systemu lub poszczególnych programów dla innych wydawnictw bibliograficznych. Wyniki realizacji ostatecznej wersji systemu pozwolą zbadać i przeanalizować jego ekonomiczną efektywność, czego nie można było dotąd dokonać w warunkach realizacji wersji próbnej.

Wydaje się również celowe opracowanie dodatkowego programu dla realizacji klasycznej postaci indeksu KWIC, co pozwoli na bezpośrednie jej skonfrontowanie z opracowaną postacią indeksu KWOC w celu ostatecznego rozstrzygnięcia teoretycznego sporu na temat przewagi jednej z tych metod w warunkach możliwie najbardziej obiektywnych, i co istotniejsze, w odniesieniu do specyfiki języka polskiego.

Realizację systemu KWOC należy uznać za dalszy konsekwentny krok w kierunku zainicjowanego przez CIINTE praktycznego opanowywania problematyki automatyzacji procesów informacyjnych, dziedziny zastosowań, która aktualnie w skali światowej wykazuje najwyższe tempo rozwoju.

681.322.004.14:681.327.004.14:002.513.5

Przegląd urządzeń do wyszukiwania i gromadzenia informacji[1)]

Tłumaczenie artykułu L. H. Berul, „Datamation", marzec 1968, nr 3. Stwierdzono opłacalność stosowania maszyn matematycznych trzeciej generacji do niektórych procesów informacyjnych, szczególnie przy pracy „on-line". Przedstawiono urządzenia komunikacji człowiek—maszyna (urządzenia końcowe, monitory ekranowe, foniczne urządzenia wyjściowe, optyczne czytniki znaków), urządzenia do przechowywania i wyszukiwania informacji w postaci obrazów (mikrofilmowe), urządzenia do powiadamiania (drukarki, aparatura reprodukcyjna, maszyny do fotoskładania, urządzenia telefotografii). Przeprowadzono ocenę kosztów tych urządzeń.

W ciągu ostatnich lat nastąpiła doniosła zmiana w sposobach rozwiązywania problemów wyszukiwania, gromadzenia i rozpowszechniania informacji bibliograficznych (*information storage and retrieval*). Do niedawna przeważały dwa sposoby.

Pierwszy z nich wymagał drogich, specjalnie do tego przystosowanych urządzeń mikrofilmowych, jak MINICARD, WALNUT, FILE SEARCH, CRIS i inne, które na ogół umożliwiały tylko automatyzację jednej lub dwóch czynności, np. przeszukiwanie lub wyszukiwanie (czy też wybieranie) dokumentów.

Drugi sposób, o którym istnieje wprawdzie obszerna literatura, ale który właściwie zrealizowany został tylko w ograniczonej mierze w systemach eksperymentalnych — stanowi próbę zautomatyzowania całego procesu gromadzenia i wyszukiwania informacji, łącznie z indeksowaniem, sporządzaniem abstraktów oraz kontrolą słownictwa.

Obie metody nie przyniosły jednak praktycznego i skutecznego rozwiązania problemu. W próbach wykorzystania maszyn matematycznych natrafiono na szereg przeszkód metodologicznych i technologicznych takich, jak np. trudności automatycznego opracowywania indeksów i abstraktów, wysokie koszty przetwarzania i przekazywania za pomocą maszyny matematycznej oraz nieadekwatność urządzeń komunikacyjnych człowiek—maszyna.

W konsekwencji zredukowano drastycznie rolę maszyny matematycznej tak, że większość czynności w systemach gromadzenia i wyszukiwania informacji trzeba było wykonywać ręcznie. Urządzenia mikrofilmowe umożliwiały rozwiązywanie tylko nieznacznej części problemu, a płynące z tego korzyści nie uzasadniały poważnego przecież wydatku rzędu 20 000 dolarów, a nawet więcej.

Wobec zaistniałych trudności, nie udało się dokonać większego postępu w kierunku uzyskania rozwiązania problemu w pełnym zakresie. Stosowanie maszyn matematycznych ograniczono do dość małej liczby zakrojonych na bardzo szeroką skalę systemów, zaś automatyczny system filmowy okazał się uzasadniony jedynie w niewielu sytuacjach. Taki stan trwał do chwili ukazania się trzeciej generacji maszyn matematycznych.

Nowe maszyny matematyczne w połączeniu z techniką mikrofilmową i innymi specjalnymi urządzeniami dają dość szeroką skalę stosunkowo mało kosztownych rozwiązań, co uzasadnia ich stosowanie do większej niż dotąd liczby czynności w systemach wyszukiwania informacji.

Jedną z głównych przyczyn trudności rozwiązania problemu gromadzenia i wyszukiwania informacji jest ogromna różnorodność zastosowań. Polegają one na ogół na sześciu podstawowych czynnościach, których trudność wykonywania i znaczenie zależą od rodzaju referowanego materiału: dane, dokumenty lub rysunki, występujące osobno lub w połączeniu. Również ekonomiczny efekt automatyzacji jest różny w zależności od zastosowania. Automatyzacja zależy bowiem nie tylko od rodzaju materiałów, którymi się operuje, lecz również od ich objętości oraz od zamierzonego sposobu ich zużytkowania.

● Pierwszą podstawową czynnością w systemach gromadzenia i wyszukiwania informacji jest w p r o w a d z a n i e materiałów (wejście).

Należy materiały przeanalizować, sprawdzając przy tym, czy nie znajdują się już w zbiorze, skatalogować je oraz sporządzić opis bibliograficzny w postaci np. cytaty, abstraktu oraz/lub szeregu symboli indeksowych.

W systemie przetwarzania danych — czynność wprowadzania ogranicza się w zasadzie do sporządzenia zapisu danych. Natomiast w systemie operującym dokumentami, czynność ta odgrywa poważną rolę, gdyż dokumenty mogą być identyfikowane według niemal niezliczonej ilości kryteriów. Każdy dokument należy skatalogować za pomocą opisu, zawierającego tytuł, autora, źródło, datę oraz wszelkie inne informacje przydatne dla ogólnego zidentyfikowania tego dokumentu. Ponadto należy sporządzić szczegółowszy opis zawartości dokumentu w postaci abstraktu oraz szeregu symboli indeksowych. Jest to typowa czynność umysłowa.

● Drugą czynnością jest g r o m a d z e n i e. W systemach obejmujących wyłącznie dane cyfrowe lub mieszane alfanumeryczne — maszynowe nośniki pamięciowe mogą być dobierane w zależności od praktycznych potrzeb. W systemach operujących dokumentami lub materiałami graficznymi stosuje się na ogół mikrofilmy.

1) Lawrence H. BERUL, „DATAMATION" nr 3/1968.

Systemy dokumentacyjne mają do czynienia z materiałem bibliograficznym. W odróżnieniu od dokumentów pierwotnych, opisy bibliograficzne nadają się do przechowywania na maszynowych nośnikach pamięciowych i rzyczywiście w wielu dużych systemach przedrukowanie kartotek bibliograficznych odbywa się za pomocą maszyn matematycznych.

● Dalsza czynność to p o d a w a n i e d o w i a d o m o ś c i użytkowników informacji o nowych dostępnych materiałach. W systemach przetwarzania danych zawiadamianie polega po prostu na sporządzaniu specjalnych wykazów lub sprawozdań takich, jak informatory lotnicze lub statystyczne. W systemach dokumentacyjnych stosuje się różnorodne środki aktualnego informowania użytkowników, począwszy od indeksów KWIC (*Key Word in Context*) — do indywidualnie skierowywanych notatek systemu selektywnego rozpowszechniania informacji (*Selective Dissemination of Information* — SDI). W każdym z tych przypadków maszyny matematyczne odgrywają poważną rolę.

● Czwarta czynność — to s z u k a n i e-o d n a j d y w a n i e materiałów potrzebnych użytkownikowi. Jest to dziedzina, w której maszyny matematyczne mogą być bardzo przydatne. Szybkość, z jaką mogą one operować danymi i kartotekami bibliograficznymi, w sposób radykalny zredukowała czas niezbędny do przeszukiwania wielkich zbiorów dokumentów. Natomiast nie daje się zautomatyzować sam proces umysłowy właściwego formowania zapytań, pomagających w odszukaniu. Zapytanie konstruuje się za pomocą znormalizowanych symboli lub terminów danego systemu. Wprawdzie jest to problem występujący w szerszej skali w systemach dokumentacyjnych aniżeli w istniejących systemach przetwarzania danych, to jednak w miarę wzrostu liczby danych gromadzonych i przetwarzanych za pomocą maszyn matematycznych oraz skomplikowania organizacji systemów przetwarzania danych — będzie on wymagał zwrócenia uwagi również i w tych systemach.

● Inna czynność — to w y s z u k i w a n i e — f i z y c z n e u d o s t ę p n i a n i e materiałów zidentyfikowanych jako odnoszących się do zapytania użytkownika.

Jest to chyba najłatwiejsza czynność w obu rodzajach systemów. Jeśli zbiór jest zapisany w pamięci maszyny matematycznej — to wówczas wyszukiwanie jest, oczywiście, całkowicie automatyczne. W przypadku zbiorów opartych na mikrofilmach, wyszukiwanie może być bądź automatyczne, bądź też ręczne, z tym że czas znajdywania odpowiedzi będzie się nieznacznie różnić. Jednostkowe zapisy mikrofilmowe, jak np. mikrofiszki, zaopatrzone są na ogół w tytuł i adres, który można odczytać gołym okiem; są one układane w przenośnych kartotekach. Dysponując adresem jakiegoś pojedynczego zapisu mikrofilmowego, można własnoręcznie wyszukać ten zapis w ciągu 10 sekund, co niemal dorównuje czasowi wykonania tej czynności przez większość zautomatyzowanych systemów mikrofilmowych.

● Ostatnia czynność — to d o s t a r c z e n i e materiałów użytkownikowi. W większości przypadków odbywa się to pocztą ze względu na wysokie koszty wszelkiej innej formy łączności.

*
* *

Przegląd powyższych czynności jasno unaocznia fakt, że uzasadniony z technicznego i ekonomicznego punktu widzenia stopień automatyzacji systemu gromadzenia i wyszukiwania informacji w dużym stopniu zależy zarówno od rodzaju materiałów, którymi się operuje, jak i od wymagań, jakie temu systemowi stawia dane zastosowanie. Panuje na ogół zgodny pogląd, że praktycznie nie można wyeliminować pracy człowieka w systemie gromadzenia i wyszukiwania informacji. Wiele czynności wymaga pracy jego umysłu, inne zaś mogą być wykonywane sposobem ręcznym, taniej niż automatycznie.

Po stwierdzeniu praktycznej niemożliwości zautomatyzowania całości procesu gromadzenia i wyszukiwania informacji oraz niewielkiej efektywności ekonomicznej specjalizowanych systemów z ograniczonymi czynnościami, czas zapoznać się z właściwą rolą maszyn matematycznych w systemach wyszukiwania informacji.

Istnieją trzy dziedziny, w których maszyny matematyczne mają znaczną rolę do odegrania, a mianowicie: przetwarzanie, współdziałanie człowiek—maszyna oraz sterowanie specjalizowanymi urządzeniami zewnętrznymi.

Maszyna matematyczna ma, oczywiście, do spełnienia szereg funkcji przetwarzających w zakresie gromadzenia danych bibliograficznych, zestawiania indeksów i innych środków powiadamiania, przeszukiwania kartotek i wydawania w drukowanej formie wyników poszukiwań. Ze względu na to, że w systemie pozostaje człowiek—maszyna matematyczna musi stwarzać możliwość nie istniejącego poprzednio wzajemnego oddziaływania maszyna—człowiek.

Ponadto — wobec istnienia urządzeń, które mogą służyć do automatyzacji różnych czynności procesu, jak np. składu drukarskiego, reprodukcji, przekazywania pytań-odpowiedzi, wybierania dokumentów a także transmisji fototelegramów, byłoby wskazane zrealizowanie sterowania tymi urządzeniami za pomocą maszyn matematycznych.

Trzecia generacja maszyn matematycznych w sposób dostateczny umożliwia realizację tych funkcji, tj. współdziałania człowiek—maszyna, programowego przetwarzania i sterowania różnymi urządzeniami.

Elastyczność, mniejsze koszty

Trzecia generacja maszyn matematycznych obniżyła koszty przetwarzania 1 miliona rozkazów z 5 dolarów w urządzeniach drugiej generacji do 0,15 dol. Niskie koszty oraz rozwój systemów wieloprogramowych umożliwiają wykorzystanie jednej maszyny do różnych czynności gromadzenia i wyszukiwania informacji względnie też do dodatkowego obciążenia tymi czynnościami maszyny, przeznaczonej do innych zastosowań. W obu przypadkach koszty zautomatyzowanego systemu wyszukiwania informacji są znacznie zredukowane.

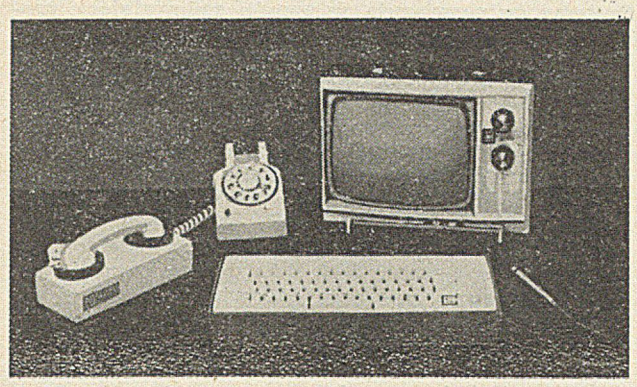
Pewne znaczenie miał również fakt zwiększenia pojemności i potanienia pamięci. Pamięci masowe z dostępem wyrywkowym osiągnęły obecnie takie rozmiary przy stosunkowo niewysokim koszcie, że umożliwiają magazynowanie wielkich zbiorów „*on-line*". Efektem tego rozwoju jest to, że maszyna matematyczna może obecnie na bieżąco obsługiwać podłączone urządzenia końcowe *(terminal)*, wykonując jednocześnie inne drugoplanowe zadania z zakresu przetwarzania. Stwarza to celowość odegrania przez maszynę matematyczną pomocniczej, ale ważnej roli w wykonywaniu szeregu czynności systemu wyszukiwania informacji, których automatyzacja była uprzednio raczej nieopłacalna.

Operacje wejścia do systemu przetwarzania danych mogą być dokonywane bezpośrednio ze sprzężonych urządzeń końcowych, o ile jest to uzasadnione właściwościami danego zastosowania. W systemach dokumentacyjnych cała operacja katalogowania może się odbywać „*on line*". Jest to o tyle korzystne, że w ten sposób maleje znacznie ilość operacji papierkowych w procesie wprowadzania informacji. Również czynności sterownicze, które zwykle pociągają za sobą ładowanie i wyładowywanie całych partii materiału w miarę ich przesuwania się z jednego stopnia na drugi, są znacznie uproszczone. Maszyna matematyczna śledzi przebieg procesu za pomocą „*on-line*" zbiorów dziennych *(transactions)*, które przetwarza. W przypadku dużego systemu zredukowanie operacji natury fizycznej oraz uproszczenie procedury sterowniczej przyczynia się w znacznej mierze do podniesienia efektywności systemu.

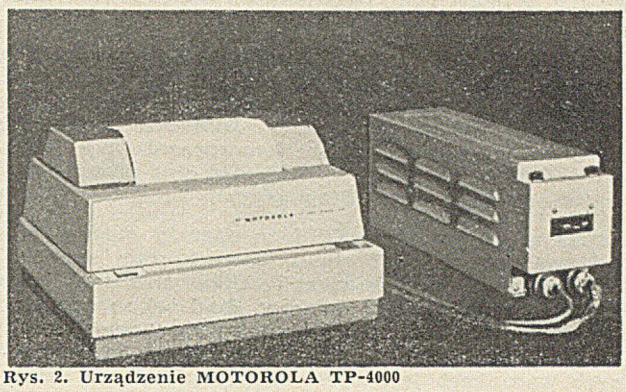
Operacja „*on-line*" wpływa również na umysłowe czynności oraz na formułowanie zapytań poszukiwawczych. Obą te zadania mogą być wykonywane szybciej i skuteczniej przy użyciu tezaurusa przechowywanego w pamięci maszyny matematycznej lub też jakiegoś innego wykazu, którego poszczególne części mogą być pokazywane na żądanie. Czas odszukiwania można nawet jeszcze bardziej skrócić, jeżeli maszyna matematyczna będzie podawała poszukującemu, jaka liczba pozycji w zbiorze odpowiada jego zapytaniu oraz będzie wskazywała próbki pomagające w sformułowaniu bardziej szczegółowych zapytań.

Urządzenia komunikacji człowiek—maszyna

Nie tylko usprawnienia w zakresie gromadzenia i wyszukiwania informacji każą nam patrzeć z optymizmem na tę dziedzinę. M.in. dokonano również szeregu istotnych ulepszeń w urządzeniach peryferyjnych. Jednymi z nich są monitory ekranowe odwzorowujące informacje w postaci znaków alfanumerycznych. Cena ich wynosi 5000 do 10 000 dolarów, natomiast koszt wynajmowania — zaledwie 125 dola-



Rys. 1. Urządzenie CC-30



Rys. 2. Urządzenie MOTOROLA TP-4000

rów miesięcznie. Urządzenia tego typu składają się ze wskaźnika ekranowego wyposażonego w elektronową lampę promieniową, manipulatora operacji graficznych oraz klawiatury, która służy do zapisu i sterowania. Działają one stosunkowo szybko, zwłaszcza w przypadku takich operacji, jak wprowadzanie i wydawanie materiału bibliograficznego oraz przy wyszukiwaniu, które należy realizować w ramach naturalnej metody konwersacyjnej z systemem. Typowymi przykładami tego rodzaju urządzeń są: IBM 2260, RCA 70/752, RAYTHEON DIDS-400, SANDERS 720, BUNKER RAMO SERIES 200, CDC 210, GE DATANET-760 oraz STROMBERG CARLSON SC 1100. Wadą tych przystawek jest niezdolność do operowania materiałem graficznym oraz — bez wyposażenia w pomocnicze dalekopisy — do transmisji danych w formie depesz — *(hard copy)*.

Do zastosowań, wymagających skromnych graficznych zdolności może służyć względnie tania przystawka CC-30 (rys. 1), produkcji firmy Computer Communications, Inc. Zestaw składający się z urządzenia sterującego, pióra świetlnego, klawiatury i przenośnego odbiornika telewizyjnego, używanego jako monitora, można nabyć za ok. 8500 dolarów. Cena bardziej skomplikowanych graficznych systemów we—wy przekracza znacznie sumę 25 000 dolarów, przy czym na ogół wymagają one specjalnych urządzeń przystosowujących je do pracy ze zdalnym kierowaniem.

W dziedzinie zastosowań monitorów ekranowych do transmisji danych w formie depesz istnieje szereg rozmaitych urządzeń o różnych szybkościach przesyłania; począwszy od szybkości pisania na maszynie 10—15 znaków/sekundę aż do 60 000 znaków/sek. Sięganie do górnej granicy szybkości jest raczej niepraktyczne, albowiem wymagałoby zastosowania bardzo drogich szerokopasmowych linii komunikacyjnych. Urządzenia wolne mają tę wadę, że nie wykorzystują w pełni linii telefonicznych pracujących w odpowiednim zakresie częstotliwości akustycznych. Problem ten można jednak rozwiązać za pomocą lokalnego sterowniczego urządzenia komunikacyjnego, obsługującego zespół urządzeń końcowych o małej szybkości, tzn. o szybkości pisania na maszynie. Zazwyczaj nie pracuje się w żadnym z tych dwóch skrajnych przypadków.

Można znaleźć szereg urządzeń, działających z szybkością 300 znaków/sekundę lub w pobliżu tej wartości, niezbędnej do pełnego wykorzystania zwykłych linii telefonicznych.

Typowym tego rodzaju urządzeniem jest MOTOROLA TP 4000 (rys. 2) model pulpitowy o wydajności 120—300 znaków/sekundę, cena nieco poniżej 10 000 dolarów. Firma Teletype Corp. zapowiedziała wypuszczenie wkrótce podobnego urządzenia o szybkości 120 znaków/sekundę, nazwanego INKTRONIC KS.

Tego rodzaju sprzęt nie jest niezbędnym uzupełnieniem monitorów ekranowych. Zapewne większość zastosowań w systemie gromadzenia i wyszukiwania informacji nie stawia wymagań co do szybkości odwzorowania i w tych przypadkach urządzenia z klawiaturą przystosowane do transmisji danych w formie depesz stanowią w pełni zadowalające i tańsze urządzenie końcowe, pracujące w reżimie *„on-line"*. O wiele tańszym od urządzenia typu *„on-line"* z monitorem ekranowym jest foniczne urządzenie wyjściowe — *(voice-response unit)*. Oczywistym plusem tego urządzenia jest fakt, że telefon, za który opłata miesięczna wynosi 5,50 dol. staje się tu urządzeniem końcowym typu *„on-line"*. Kosztuje to — rzecz jasna — 500 dolarów miesięcznie. Ponieważ jedno urządzenie może w tym samym czasie — w zależności od zestawu — obsługiwać mniej więcej od 4 do 48 linii, to oczywiście koszt przypadający na poszczególnego użytkownika jest raczej niski. Istnieje wiele urządzeń tego typu. Firma IBM dostarcza urządzenie 7770 o słowniku zawierającym 128 wyrażeń, jak również urządzenie 7772 z utrwalonym słownikiem o 1000 wyrazach, mogące jednocześnie syntetyzować dodatkowe wyrazy z podstawowych jednostek mowy, zapamiętanych przez to urządzenie.

Firma RCA wyprodukowała urządzenie 70/510, którego maksymalny zasób wynosi 189 wyrazów. COGNITRONICS dostarcza linię z urządzeniami, zwaną SPEECHMAKER, którą można podłączyć do niemal każdej maszyny matematycznej.

Jedną z wad fonicznego urządzenia wyjściowego jest ograniczoność słownika. IBM rozwiązała ten problem w swoim urządzeniu 7772 przez zapisanie w pamięci podstawowych dźwięków w postaci kodów cyfrowych, a nie przez wstępny zapis całych wyrazów. Dźwięki mogą być składane za pomocą syntetyzującego „urządzenia kodującego dźwięk" *(vocoder) tak*, aby można było uzyskać dość znaczną ilość różnorodnych wyrazów.

Inny sposób zwiększania rozmiarów słownika może być zrealizowany przez powiększenie pamięci słów. Niezależnie jednak od rozmiarów słownika, urządzenie foniczne przydatne jest tylko w tych rodzajach systemów danych, których celem jest dostarczanie odpowiedzi na bardzo specyficzne zapytania na podstawie danych z kartoteki o wielkiej ilości dokumentów. W wielu tego rodzaju zastosowaniach — urządzenie takie przynosi dobre rezultaty.

Najbardziej znany jest chyba system amerykańskiej giełdy, która posługuje się urządzeniem COGNITRONICS SPEECHMAKER w celu dostarczania notowań giełdowych.

Inny system — to EDICT, używany przez oddział systemów informacji zakładów NORTH AMERICAN ROCKWELL w celu dostarczania danych konstrukcyjnych dla personelu technicznego. Dane z liczby ponad 75 000 rysunków technicznych można uzyskać z każdego z 42 000 telefonów tych zakładów za pośrednictwem urządzenia IBM 7770.

Optyczne czytniki znaków

Inne ważne urządzenia peryferyjne, których wpływ w dziedzinie gromadzenia i wyszukiwania informacji jest na razie ograniczony, lecz będzie stale wzrastał, to optyczne czytniki znaków. Urządzenia te,

zwane urządzeniami optycznego rozpoznawania znaków (ORZ) dzielą się na dwie grupy.

Pierwsza — to czytniki mogące rozpoznawać tylko specjalne pisma stylizowane, jak np. urządzenie IBM 1428 oraz FARRINGTON model 3010 Optical Scanner. Typ pierwszy czyta pismo IBM 1428, natomiast drugi — pisma stylizowane: samosprawdzające FARRINGTON SELF-CHECK, IBM 1428 oraz pismo opracowane przez Instytut Standaryzacji USA.

Druga grupa urządzeń ORZ to aparaty mogące odczytywać pewną ilość standardowych pism stosowanych w maszynach do pisania. Urządzenia tego typu kosztują ponad 500 000 dolarów. Stąd też są one stosowane dopiero wtedy, gdy przeważy argument wielkiej ilości materiału do odczytania, który w normalnych warunkach jest wprowadzany za pomocą dziurkowania taśmy lub kart.

Inną ich wadą jest konieczność uprzedniej identyfikacji rodzaju pisma, które maszyna ma rozpoznać.

Typowy dla tych czytników jest ELECTRONIC RETINA COMPUTING READER (Siatkowy Elektroniczny Czytnik Obliczeniowy), produkcji firmy RECOGNITION EQUIPMENT, Inc., który w praktyce może rozpoznawać niemal każdy na stałe wprowadzony typ pisma, z ręcznym pismem drukowanym łącznie. Inne urządzenia tej klasy to PHILCO-FORD GENERAL PURPOSE PRINT READER, mogący rozpoznawać każde z ośmiu podstawowych pism maszyn do pisania, jak również inne odmiany pism. Należy także wymienić IBM 1975 OPTICAL PAGE READER, mogący rozpoznawać do 200 rodzajów pisma.

Ze względu na istniejące problemy wielu rodzajów pism oraz koszty, stosowanie tych maszyn w systemach gromadzenia i wyszukiwania informacji jest na razie ograniczone do wprowadzania umiarkowanie wielkich ilości zapisów drukowanych lub wykonanych ręcznym pismem drukowanym. Na wykorzystanie ORZ do celów wprowadzania drukowanej literatury publikowanej — trzeba będzie zaczekać do chwili osiągnięcia większej elastyczności tego systemu.

Przechowywanie i wyszukiwanie informacji w postaci obrazów

Szereg różnorodnych specjalizowanych urządzeń, umożliwiających usprawnienie czynności gromadzenia, wyszukiwania, powiadamiania i udostępniania informacji, zwłaszcza w zastosowaniu do systemów operujących dokumentami i materiałem graficznym — stanowi doskonałe uzupełnienie zwiększonych możliwości maszyn matematycznych trzeciej generacji i ich urządzeń peryferyjnych.
Sprzęt służący do gromadzenia i wyszukiwania składa się z urządzeń do mikrofilmowania dokumentów pierwotnych oraz do wyszukiwania dokumentów ze zbiorów opartych na mikrofilmach. Aparaty do filmowania dokumentów dostarczają firmy RECORDAK BELL i HOWELL, IBM, XEROX, REMINGTON RAND, STROMBERG-CARLSON, 3M, NCR, PHOTO DEVICES i inne.

Dostępne są również różnorodne urządzenia do uzyskiwania z rolek filmowych jednostkowych form zapisu w postaci kart aperturowych (z okienkiem) mikrofiszek [2]) lub zasobników.

W dziedzinie przetwarzania mikrofilmów na dokumenty nie osiągnięto jeszcze zadowalających postępów. Istnieje szereg czytników-drukarek, jak np. ITEK 1824 lub 3N FILMAC 200, które pod kontrolą operatora mogą wykonywać odbitki z każdego rodzaju mikrofilmu. Problem polega jednak na tym, że proces wykonywania odbitek jest zbyt powolny, a koszty — zbyt wysokie. Jedyną znaną wydajną maszyną, produkującą dokumenty z mikrofilmów jest XEROX COPYFLO. Dokumenty te produkuje ona jednak jedynie z rolek taśmy filmowej.

Istnieje szereg systemów, które uniknęły błędów pierwszych mikrofilmowych systemów wyszukiwania. Systemy te rozwiązują w sposób zadowalający problem kosztów—wydajności łącząc pracę systemem ręcznym z wysoce zautomatyzowanym. Większość tego rodzaju systemów wykorzystuje formy zapisów jednostkowych, jak np. mikrofiszki oraz stosuje sposób odszukiwania informacji raczej za pomocą bezpośrednich adresów aniżeli opisów bibliograficznych. Jednym z nich — to produkowany przez HOUSTON FEARLESS CORP. system COMPACT AUTOMATIC RETRIEVAL DISPLAY (CARD) — Monitor Ekranowy do Automatycznego Wyszukiwania. Cena jego kształtuje się poniżej 5000 dolarów, pojemność wynosi 73 500 stron, zaś czas odpowiedzi — mniej niż 5 sekund. Adres dokumentu wprowadzany jest za pomocą klawiatury; można go jednak przystosować do przyjmowania taśmy papierowej czy też sygnałów cyfrowych bezpośrednio z maszyny matematycznej.

Systemem o większej pojemności, operującym kartami dziurkowanymi lub mikrofiszkami jest SELECTRIEVER, produkowany przez MOSLER SAFE Comp. Cena tego systemu sięga ponad 20 000 dolarów. Posiada on zdolność zmagazynowania 200 000 jednostek mikrofilmowych oraz umożliwia odszukiwanie w ciągu 6 sekund. Wyszukiwanie odbywa się za pomocą liczb ewidencyjnych *(accession numbers)*, które mogą być wprowadzone bądź za pomocą klawiatury, bądź z taśmy papierowej.

Bardziej złożona wersja tego podstawowego systemu umożliwia automatyczne wyświetlanie na monitorze ekranowym i powielanie wyszukanego materiału oraz jego transmisję do innego odległego ekranu.

Najbardziej udanym spośród systemów operujących rolkami taśmy mikrofilmu jest system MIRACODE RECORDAK, łączący czynność odszukiwania z wyszukaniem dokumentu. Używa on kaset zawierających 16-milimetrowy mikrofilm o długości 100 stóp. Każda kaseta zawiera 2000 klatek, z których każda może być odszukana za pomocą danej cechy lub cech *(attribute)* w ciągu 10 sekund, rzutowana na ekran lub też skopiowana. W przypadku konieczności przebadania szeregu kaset w celu znalezienia odpowiedzi na dane zapytanie, czas wyszukiwania jest, oczywiście, dłuższy. System podstawowy, składający się z kamery i jednostki wyszukującej kosztuje ok. 30 000 dolarów.

Aparatura do powiadamiania

Urządzenia do powiadamiania dzielą się również na dwie klasy. Jedna z nich składa się z urządzeń do wytwarzania taniej, częstej, bieżącej informacji o nowych, aktualnych materiałach. Ze względu na to, że w bieżącym powiadamianiu o aktualiach szybkie działanie odgrywa zasadniczą rolę, jakość produkcji dokumentów ustępuje wymaganiom szybkości.

Najprostszym istniejącym urządzeniem tego rodzaju powiadamiania jest szybka komputerowa drukarka łańcuchowa o repertuarze, obejmującym duże i małe litery. Klasa urządzeń o bardziej skomplikowanej konstrukcji służy przetwarzaniu materiału z taśmy magnetycznej na mikrofilm. Odbywa się to przez wytwarzanie znaków na powierzchni elektronowej lampy promieniowej i kopiowanie ich na mikrofilmie.

Firma STROMBERG-CARLSON dysponuje szeregiem tego rodzaju urządzeń do zastosowań operujących małą, średnią i dużą ilością materiału. Rejestrator SC 4460 tej firmy przekształca materiał zapisany na taśmie magnetycznej na zapis na rolce filmowej lub mikrofiszkach z szybkością sięgającą 90 000 znaków/sek. Znaki są rozmieszczane proporcjonalnie, a nadruki składane przez maszynę matematyczną mogą być wprowadzane na mikrofilm, co pozwala wyeliminować koszty oddzielnego wprowadzania nadruków.
Firma 3M oferuje dodatkowe ulepszenie w swoich urządzeniach Series F ELECTRON BEAM RECORDER, polegające na eliminacji wilgotnego wywoływania, normalnie stosowanego przy wytwarzaniu materiału mikrofilmowego. Uzyskuje się w ten sposób szybsze przetwarzanie.

[2]) Mikrofilmy w postaci arkuszy o małych wymiarach.

Firma XEROX oferuje dwie maszyny, nadające drukowanym materiałom otrzymanym z maszyny matematycznej postać znacznie bardziej odpowiadającą celom powiadamiania. Jedna z nich — to XEROX 2400-IV, która redukuje rozmiary druków komputerowych z 11″ × 17″ do 8½″ × 11″. W zastosowaniach wymagających dużej ilości informacji (przeszło 200 000 odbitek miesięcznie), koszty jednej odbitki spadają niemal do 1 centa amerykańskiego. W porównaniu z drukiem offsetowym — maszyna 2400-IV szczególnie nadaje się do produkcji małych ilości odbitek, sięgających 50 lub mniej egzemplarzy. Można tu posługiwać się przezroczystymi matrycami, zawierającymi standardowe nadruki, eliminując operację wcześniejszego ich drukowania (nadruków).

Druga maszyna firmy XEROX — to urządzenie 3-2-1, posługujące się półmikrokserografią (SMX) do uzyskania 3-krotnego zmniejszenia obrazu na kartach. Typowym formatem wyjściowym jest obraz na karcie informacyjnej (tabulogramie), na której można wydziurkować inne informacje za pomocą klawiatury lub nanieść na brzegach znaki do celów przetwarzania maszynowego. Ze względu na to, że rozmiary obrazu pozwalają na odczytywanie informacji gołym okiem, karta może być wykorzystana jako powiadomienie oraz jako dokument z danymi, które mogą być przetwarzane przez maszynę matematyczną. W urządzeniu tym można stosować także karty o innych rozmiarach, aż do 8″ × 8″.

Firma NCR osiągnęła ostatnio nowy rekord w dziedzinie mikroobrazkowej, produkując PHOTO-CHROMIC MICRO-IMAGE SYSTEM (Fotochromowy Mikroobrazkowy System PCMI), mogący sprowadzać 3200 stron do wielkości pojedynczej znormalizowanej mikrofiszki o wymiarach 4″ × 6″ z szybkością 1000 stron na godzinę. Dzięki znacznej skali pomniejszenia (200 razy) koszty odbitki są bardzo niskie. Kopia mikrofiszki 4″ × 6″, zawierająca 3200 obrazów może być wyprodukowana po cenie od 50 centów do 1 dolara w zależności od ilości. Aparaty do rzutowania można wynajmować za 10—15 dolarów miesięcznie; mogą one być zaopatrzone w przystawki służące do wykonywania dokumentów.

Przewiduje się, że urządzenia PCMI będą miały zastosowanie przede wszystkim w tych dziedzinach, w których informacja podlega częstej aktualizacji, jak np. w przypadku katalogów i indeksów. Koszty odbitek są na tyle niskie, że opłaca się raczej zastąpić nowym — stary katalog lub indeks, aniżeli aktualizować go. Przy cenie 1 dolara za mikrofiszkę, zawierającą 3200 obrazów (stron), nowy katalog o objętości 10 000 stron może być sporządzony za 3 dolary. Jest to różnica ok. 30 dolarów w porównaniu z techniką wytwarzania mikrofiszek o normalnym 15—18-krotnym zmniejszeniu.

Inna klasa urządzeń do powiadamiania pozwala produkować w formie książkowej indeksy i katalogi, publikowane w dłuższych — półrocznych lub rocznych cyklach. Tego rodzaju materiał z dziedziny powiadamiania odznacza się wyższą jakością, szatą graficzną, tzn. różnorodnymi typami pisma drukarskiego, nie zniekształconymi i proporcjonalnie rozmieszczonymi znakami oraz gęstością pionową 250—300 wierszy na cal.

Największym osiągnięciem na tym polu jest fotoskładanie.

Jedną z pierwszych maszyn fotoskładających, skonstruowanych we Francji jest PHOTON. Maszyna ta może wyprodukować gotowe odbitki fotograficzne w 16 różnych rodzajach pisma drukarskiego i w 12 rozmiarach pola znaku rozłożonego na 5 do 95 elementów, z szybkością 8—10 znaków/sek. Wprowadzenie materiału może się odbywać bądź z klawiatury, bądź z taśmy papierowej. Nowszy model tego urządzenia PHOTON 900 jest w stanie pracować z szybkością 250 znaków/sek, przy ograniczonym repertuarze pisma.

Nowym osiągnięciem w dziedzinie fotoskładania jest MERGENTHALER LINOTRON, działający z szybkością 1000 znaków/sek. Specjalny model, skonstruowany dla wojskowego lotnictwa amerykańskiego (*US Air Force*), zwany LEXICOL-GRAPHICAL COMPOSER-PRINTER, może łączyć tekst z fotografiami i rysunkami, utrwalonymi na taśmie telerecordingu.

Firma RCA wprowadziła z dużym powodzeniem na rynek swój VIDEO-COMP-SYSTEM, który współpracuje z maszyną matematyczną SPECTRA 70 i może składać z szybkością do 600 znaków/sek.

Również firma HARRIS INTERTYPE zapowiedziała ostatnio wypuszczenie maszyny dającej 1000 znaków/sek. W sprzedaży znajduje się także wiele innych tego rodzaju urządzeń.

Wreszcie — ostatnia grupa specjalizowanych urządzeń — to urządzenia telefotograficzne do przesyłania materiału graficznego do odległych miejscowości. Telefotografia o małej szybkości przesyłania używana jest od lat przez agencje prasowe. Wersję do celów gospodarczych oferuje firma XEROX przy opłacie za wynajmowanie 35 dolarów miesięcznie. Urządzenie to, zwane MAGNAVOX TELECOPIER wymaga ok. 6 minut do przekazania materiału, umieszczonego na stronie o rozmiarach 8½″ × 11″ drogą zwykłych linii telefonicznych pracujących na częstotliwościach akustycznych.

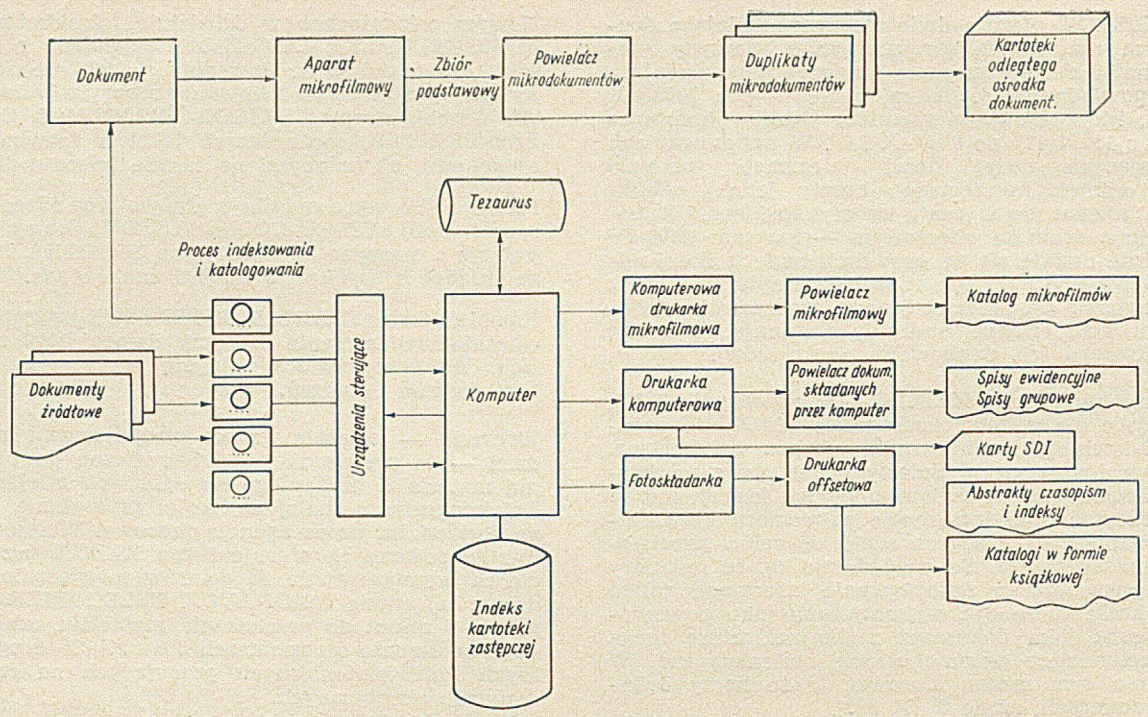
Firma MUIRHEAD and COMPANY, Ltd., oferuje szybsze urządzenie, zwane MUFAX COURIER 500, mogące przekazać stronę o rozmiarach 8½″ × 11″ w ciągu ok. 2 minut.

Firma ALDEN ELECTRONIC and IMPULSE RECORDING EQUIPMENT skonstruowała szereg urządzeń telefotograficznych, które można podłączyć bezpośrednio do zautomatyzowanego mikrofilmowego systemu wyszukiwania informacji. Szybkość transmisji w tym systemie wynosi od 4 do ½ minuty dla jednej strony w zależności od rodzaju użytych kanałów łączności.

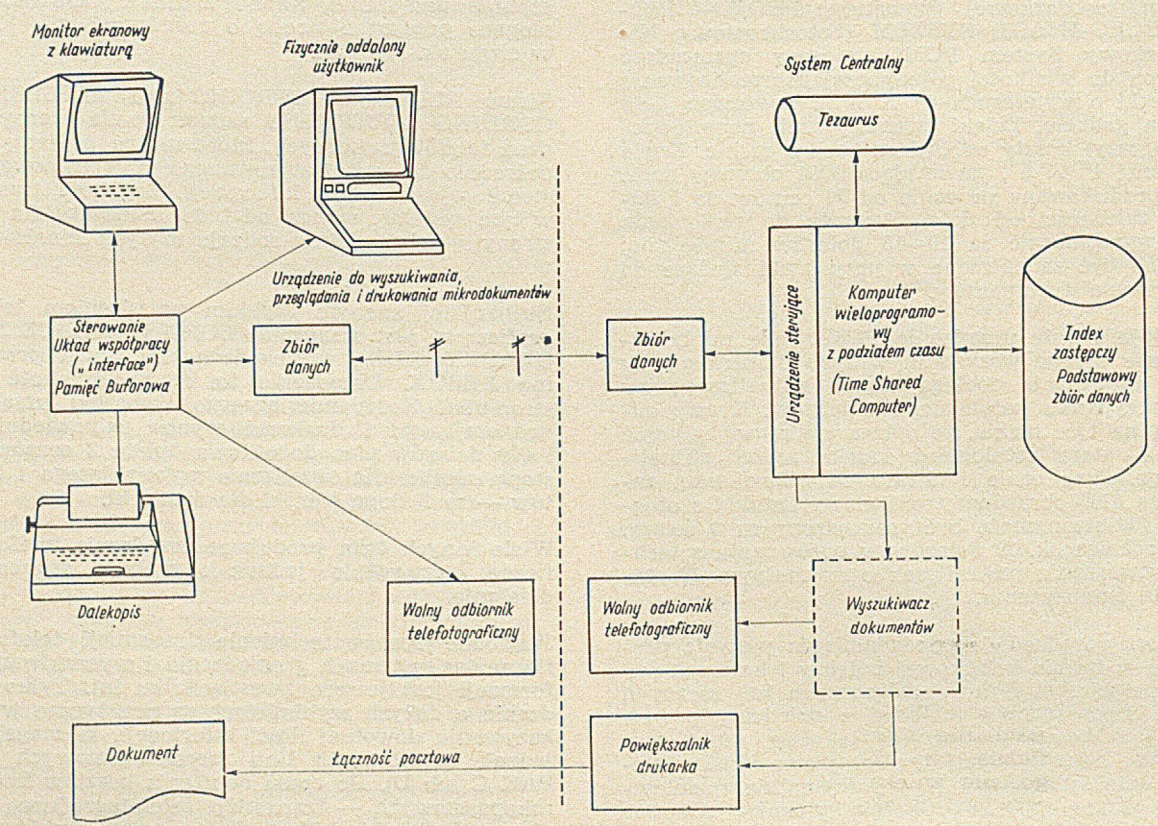
Najbardziej znanym szybkim urządzeniem telefotograficznym jest urządzenie LONG DISTANCE XEROGRAPHY (LDX), wprowadzone na rynek przez firmę XEROX. Urządzenie to osiąga szybkość do 8 stron/min. Koszt miesięcznego wynajmu urządzenia wybierającego i drukarki wynosi odpowiednio 550 i 650 dolarów plus dodatkowa opłata 2 centów za 1 stopę odbitki dla urządzenia wybierającego i 3 centów — za 1 stopę odbitki dla drukarki.

Wiele innych firm produkuje urządzenia telefotograficzne, przeważnie jednak o małych szybkościach działania.

Względne rozpowszechnienie transmisji telefotograficznej w systemach gromadzenia i wyszukiwania informacji jest jeszcze kwestią kilku lat. Tylko urządzenia o dużych szybkościach są praktyczne w przekazywaniu dowolnej ilości informacji. Wymagają one jednak kosztownych linii szerokopasmowych (TELPAK C lub D). Do czasu obniżenia kosztów linii szerokopasmowych — transmisja telefotograficzna ograniczona będzie do kilku zaledwie zastosowań, w których ilość obrazów do przesyłania jest zbyt mała, a czas na udzielania odpowiedzi na zapytania — dość długi, aby można było korzystać z wolniejszych urządzeń i tańszych kanałów łączności. Obiecującym przedsięwzięciem w tej dziedzinie jest WESTERN UNION's BROADBAND SWITCHING NETWORK (Szerokopasmowa Sieć Przełącznikowa spółki Kolejowej WESTERN UNION, która stanowi wybierany za pomocą tarczy aparatu telefonicznego szerokopasmowy system, dostępny na zasadzie opłaty za rozmowę międzymiastową. Obsługuje on 39 dużych miast. System ten pozwala uniknąć wysokich kosztów za dzierżawienie mikrofalowych radiołączy czy też linii szerokopasmowych do potrzeb przekazywania danych handlowych.



Rys. 3. Wprowadzanie i powiadamianie



Rys. 4. Zdalny system komunikacyjny człowiek-maszyna

System „state of the art" [3]

Przy wykorzystaniu istniejących obecnie maszyn matematycznych i urządzeń specjalizowanych możliwe jest zrealizowanie komunikacji człowiek—maszyna, która zwiększyłaby sprawność i efektywność operowania zbiorami materiału bibliograficznego. Rys. 3 i rys. 4 przedstawiają układ i wykorzystanie takiego systemu.

Rys. 3 odnosi się do czynności wprowadzania i powiadamiania. Nowy materiał w celu powielenia zostaje przeglądany, a następnie podlega skatalogowaniu i zaindeksowaniu za pomocą urządzenia końcowego z monitorem ekranowym, pracującego w reżimie „*On-line*" z tym, że człowiek wykonuje tu czynności wymagające pracy umysłowej, zaś maszyna matematyczna — niezbędne czynności przeszukania kartotek i wprowadzania. Ponadto maszyna matematyczna dostarcza człowiekowi działy tezaurusa lub danego wykazu źródłowego, niezbędne do wypełnienia jego czynności. Następuje mikrofilmowanie i powielenie dokumentu pierwotnego do potrzeb fizycznie odległych ośrodków.

3) System odpowiadający obecnemu stanowi ETO.

EMC wykorzystuje kartoteki bibliograficzne w celu wytworzenia szeregu różnorodnych materiałów z dziedziny powiadamiania. Drukarka mikrofilmowa sporządza katalog mikrofilmowy. Notatki z zakresu selektywnego rozpowszechniania informacji sporządza komputerowa drukarka łańcuchowa o repertuarze obejmującym duże i małe litery. Inne powiadomienia o ograniczonym rozpowszechnianiu, jak np. spisy ewidencyjne *(accession lists)* przechodzą również przez drukarkę komputerową, a następnie są reprodukowane w odpowiedniej ilości w znormalizowanych rozmiarach 8½" × 11" przez powielacz dokumentów składanych i justowanych przez EMC, jak np. XEROX 2400-IV. Abstrakty czasopism, indeksy i katalogi książkowe są składane na sterowanej przez EMC maszynie fotoskładającej, a następnie reprodukowane na maszynie offsetowej.

Rys. 4 wskazuje, w jaki sposób system ten obsługiwałby odległego fizycznie użytkownika w zakresie wyszukiwania i przesyłania.

Użytkownik formułuje swoje zapytania za pomocą klawiatury zdalnego monitora ekranowego, pracującego w reżimie „*on-line*". System może mu służyć pomocą w formułowaniu kwerendy metodą postawienia szeregu pytań oraz przez wyświetlenie odnośnych części tezaurusa zmagazynowanego w pamięci EMC.

Jeżeli miejscowa kartoteka nie zawiera pożądanego materiału dokumentacyjnego, prośba użytkownika przekazywana jest bezpośrednio do centralnej maszyny matematycznej, która wysyła duplikat oryginału bądź pocztą, bądź za pośrednictwem wolnowybierającego telefotografu. Telefotograficzny sposób dostawy trwa 5 minut i kosztuje 25 centów od strony plus opłata za połączenie międzymiastowe.

Jeżeli użytkownikowi zależy jedynie na drukowanej bibliografii, wówczas prosi o wydrukowanie cytaty. Tę cytatę otrzymuje na lokalnym dalekopisie z danych zapisanych w pamięci buforowej urządzenia sterującego.

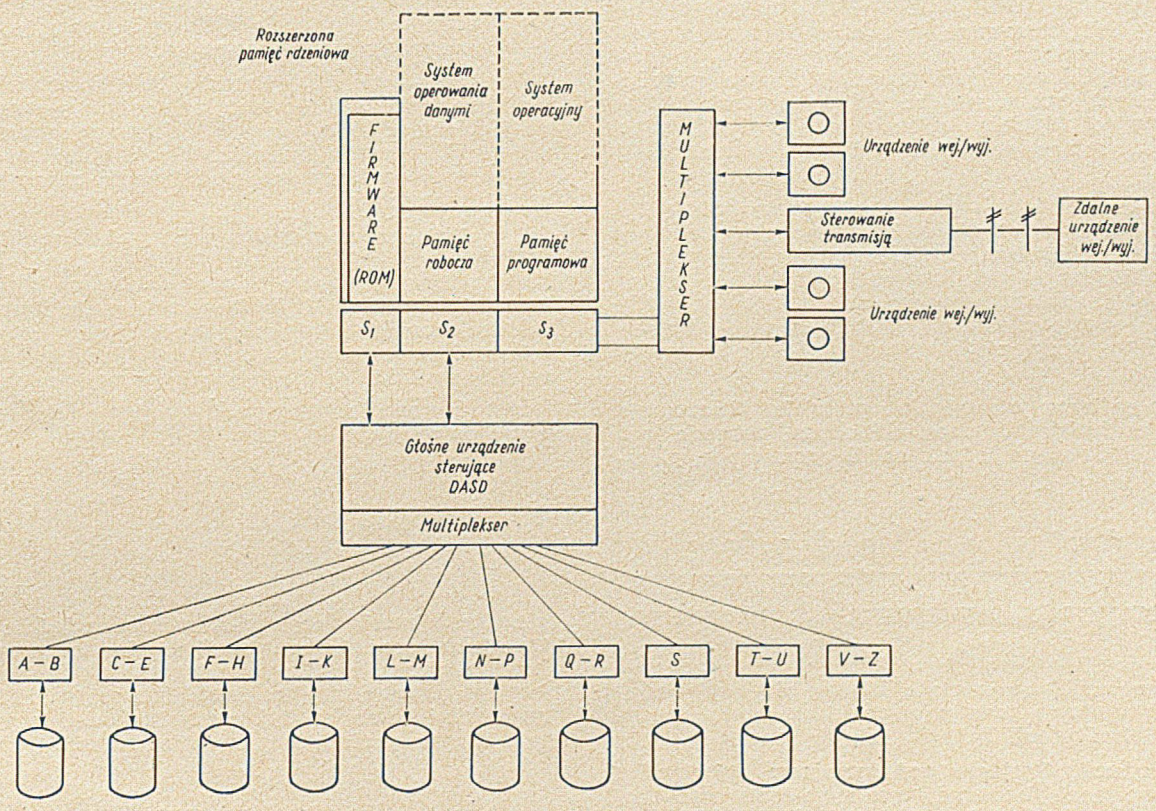
Przewidywania na najbliższe pięciolecie

Analiza kierunków rozwojowych opisanych tu urządzeń pozwala przewidzieć rozwój typowego systemu w ciągu mniej więcej przyszłych pięciu lat. Wybór urządzeń dostępnych projektantom systemów gromadzenia i wyszukiwania informacji nie ulegnie radykalnej zmianie. Klawiaturowe dziurkarki pozostaną nadal najczęściej używanymi do wprowadzania danych („*off-line*"), a to ze względu na wielkość obecnych nakładów na rozwój urządzeń, operujących kartami. Nie mniej — maszyny do pisania, które jednocześnie dziurkują taśmę papierową będą ilościowo wypierane przez maszyny do pisania z rejestratorem magnetycznym.

Optyczne czytniki znaków będą nadal użyteczne tylko do celów wprowadzania maszynopisów i materiału drukowanego przy użyciu ograniczonej ilości rodzajów kroju czcionki. Urządzenia stenodrukowe mogą wypierać w przyszłości dziurkarki klawiaturowe lub maszyny do pisania przy sporządzaniu materiału wejściowego dla czytników optycznych.

Dokończenie na IV str. okł.

SCHEMAT PRZYSZŁEJ ORGANIZACJI SYSTEMÓW KOMPUTEROWYCH



Rys. 5. Schemat przyszłej organizacji systemów komputerowych

Odpowiedź systemu na kwerendę stanowi szereg odnoszących się do danego zagadnienia elementów bibliograficznych, jak np. cytat obejmujących adres źródła dokumentów w kartotece mikrofilmowej. Użytkownik dokonuje oceny cytat i wybiera pożądany materiał.

Ponieważ system ten zakłada przekazanie pewnych duplikatów centralnego zbioru do fizycznie oddalonych użytkowników, wybór materiału dokonany przez użytkowników może być dostarczany do miejscowego wyszukiwacza-rzutnika mikrofiszek, na którym wybrany materiał ukazywałby się w czasie ok. 4 sek.

Dokończenie z III str. okł.

Urządzenia końcowe „*on-line*" znajdą znacznie szersze zastosowanie z racji istotnych korzyści ekonomicznych, jakie daje wykonywanie przez maszynę matematyczną czynności redagowania i składania. Koszty pojedynczego urządzenia końcowego, wyposażonego w promieniową lampę katodową spadną prawdopodobnie do rzędu 1000—2000 dolarów dla urządzenia pracującego w zestawie sześciu do dziesięciu sztuk, które korzystają ze wspólnego generatora znaków, urządzenia sterującego i wspólnej pamięci buforowej. Koszty te mogą być obniżone przez zastosowanie lamp z pamięcią do bezpośredniego oglądania, które nie wymagają buforowania dzięki własnościom pamięciowym względnie dzięki użyciu standardowych odbiorników telewizyjnych.

Mogą powstać możliwości zastosowań fonicznych urządzeń wejściowych, lecz w bardzo ograniczonym zakresie. Foniczne urządzenia wyjściowe będą bardziej rozpowszechnione, lecz ich stosowanie będzie ograniczone do dziedziny systemów operujących wysoko standaryzowanym materiałem. Również będzie wzrastać zakres stosowania mikrofilmów, zarówno jako środka gromadzenia informacji, jak i ich publikacji. Jednostkowe formy zapisu na mikrofilmie takie, jak karty aperturowe i mikrofiszki będą nadal najbardziej popularną postacią dokumentów. Mikrofilmy o bardzo wysokim stopniu zmniejszenia, jak np. PHOTO CHROMIC MIKRO-IMAGE (System firmy NCR) będą stosowane do publikacji często aktualizowanych materiałów, a w szczególności do sporządzania katalogów i indeksów. Przenośne rzutniki mikrofilmowe prawdopodobnie będą do nabycia w cenie poniżej 50 dolarów, co ułatwi wykorzystanie mikrofilmów jako „pakietów" (plików) informacji.

Ponadto — nastąpi większe rozpowszechnienie stosunkowo niedrogich — sięgających sumy 1000—5000 dolarów automatycznych mikrofilmowych systemów wyszukiwania informacji. Systemy te zajmować się będą tylko wyszukiwaniem mikrofilmów za pomocą adresów, zaś systemy droższe, bardziej skomplikowane, łączące czynności bibliograficznego odszukiwania i wyszukiwania — prawdopodobnie będą zanikać.

Telefotograficzne przesyłanie materiału obrazowego stosowane będzie w nieco większym zakresie niż obecnie. Będzie ono jednak nadal ograniczone, głównie do zastosowań, w ramach których będzie można korzystać z transmisji o niewielkiej szybkości za pośrednictwem akustycznych linii telefonicznych.

Największe zmiany nastąpią w organizacji centralnego ośrodka komputerowego. Za pięć lat główne człony w pełni zautomatyzowanych systemów gromadzenia i wyszukiwania informacji mogą wyglądać tak, jak to przedstawiono na rys. 5.

Będą się one różniły od obecnych pod trzema zasadniczymi względami:

Po pierwsze — „*firmware*" — termin opublikowany przez Aschera Oplera w artykule pt. „Czwarta generacja software'u", który ukazał się w „DATAMATION" ze stycznia 1967 r., używany będzie do przystosowania właściwości maszyny matematycznej do funkcji gromadzenia i wyszukiwania informacji. Będzie to oznaczało rozbudowę obecnych technik mikroprogramowania, w których maszynowe zestawy rozkazów składają się z mikro-rozkazów, zmagazynowanych w bardzo szybkiej i stosunkowo niedrogiej pamięci stałej. Przez zmianę mikro-programów możliwe będzie uzyskanie większej sprawności EMC w zakresie odszukiwania oraz przetwarzania tekstów.

Po drugie — zastosowane zostaną specjalne konstrukcje w celu zwiększenia pojemności pamięci rdzeniowej lub innego typu pamięci wewnętrznej do 20 000 000 bitów, przy koszcie 2—10 centów za bit. Stworzy to wielkie kompleksowe systemy operowania danymi oraz wielkie systemy operacyjne, które będą sprawniejsze aniżeli obecnie.

Po trzecie — dane lub kartoteki bibliograficzne gromadzone będą w masowych pamięciach o wyrywkowym dostępie, które mogą być dzielone na segmenty na zasadzie częstości użycia. Umożliwi to szeregowi użytkowników jednoczesny dostęp do różnych segmentów kartotek.

W wyniku wprowadzenia wyżej wymienionych trzech zmian — nastąpi wydatne obniżenie kosztów automatyzacji gromadzenia i wyszukiwania informacji. Dlatego też rozwój, który zapoczątkowała trzecia generacja urządzeń komputerowych będzie się nadal odbywać w przyspieszonym tempie i nie będzie żadnych przeszkód, aby każdy mógł w wysoce aktywny sposób operować zbiorami wszelkich typów i rozmiarów.

Tłum. **J. G.**

Listy do Redakcji

INFORMATYKA — INFORMIKA — INFOTRONIKA

Na marginesie artykułu doc. Marczyńskiego [1] chciałbym przytoczyć pewną może dziwną, ale bardzo użyteczną propozycję terminologiczną, zasłyszaną w kuluarach Konferencji Zakopiańskiej.

Jak wiadomo, Rosjanie pod terminem *informatika* rozumieją teorię informacji naukowo-techniczno-ekonomicznej, natomiast Francuzi pod słowem *informatique* rozumieją teorię techniki obliczeniowej w jej możliwie najszerszym rozumieniu. Te dwa podejścia były właśnie powodem pewnej kontrowersji na temat znaczenia słowa *informatyka* w języku polskim, jaka zaistniała pomiędzy referentami w trakcie wspomnianej konferencji.

Otóż zasłyszana propozycja sprowadza się do traktowania *informatyki* jako nazwy zbiorczej dla obu teorii oraz wymyślenia dla każdej z nich nowego terminu — np. można by spróbować tutaj użyć neologizmów *informika* oraz *infotronika* (względnie *informatoryka* oraz *informatronika*, ale te propozycje brzmią trochę za długo). Obrazowo propozycję tę, wraz z ukutymi ad hoc neologizmami angielskimi i rosyjskimi (inne są już oczywiste) można wyrazić przedstawionym na rysunku schematem:

INFORMATYKA

Informika informics ИНФОРМИКА	Infotronika infotronics ИНФОТРОНИКА

Muszę niestety, *relata referro*, przyznać, że przyjęcie takiej czy podobnej propozycji prowadziłoby do utworzenia polskiego odpowiednika wyrazu *computer*, którego fonetycznie spolszczona forma przyprawia niektórych o dreszcze niewytłumaczalnego obrzydzenia — postaci *infotron*, czy coś w tym rodzaju.

ABE

1) R. Marczyński — „Informatyka, czyli maszyny matematyczne i przetwarzanie informacji", „Maszyny Matematyczne" nr 1/69.